

**DETERMINATION OF
NUCLEOTIDE SEQUENCES OF THE
BACTERIOPHAGE Q β GENOME:
ORGANIZATION AND EVOLUTION
OF AN RNA VIRUS**

INAUGURAL-DISSERTATION

ZUR

ERLANGUNG DER PHILOSOPHISCHEN DOKTORWÜRDE

VORGELEGT DER

PHILOSOPHISCHEN FAKULTÄT II

DER

UNIVERSITÄT ZÜRICH

VON

PHILIPP MEKLER

VON ZÜRICH

BEGUTACHTET VON

HERRN PROF. DR. M. A. BILLETER

ZÜRICH 1981



Digitized by the Internet Archive
in 2026

**DETERMINATION OF
NUCLEOTIDE SEQUENCES OF THE
BACTERIOPHAGE Q β GENOME:
ORGANIZATION AND EVOLUTION
OF AN RNA VIRUS**

INAUGURAL-DISSERTATION

ZUR

ERLANGUNG DER PHILOSOPHISCHEN DOKTORWÜRDE

VORGELEGT DER

PHILOSOPHISCHEN FAKULTÄT II

DER

UNIVERSITÄT ZÜRICH

VON

PHILIPP MEKLER

VON ZÜRICH

BEGUTACHTET VON

HERRN PROF. DR. M. A. BILLETER

ZÜRICH 1981

DETERMINATION OF
NUCLEOTIDE SEQUENCES OF THE
BACTERIOPHAGE ϕ 8 GENOME:
ORGANIZATION AND EVOLUTION
OF AN RNA VIRUS

INHAUS JAL DISSERTATION

1988

VERGLEICHENDE GENOMANALYSE VON BAKTERIOPHAGEN UND
VIRUSGENOMEN

VERGLEICHENDE GENOMANALYSE

VERGLEICHENDE GENOMANALYSE

1988

UNIVERSITÄT ZÜRICH

1988

PHILIPPE KLEIN



UNIVERSITÄT ZÜRICH

VERGLEICHENDE GENOMANALYSE

to my parents

to TTM and Richard

My most sincere thanks go to Prof.Dr.M.A.Billeter and to Prof.Dr.C.Weissmann, under whose supervision this work gained shape, for all the support and the encouragement I was given and for the many valuable discussions I was having the privilege to enjoy.

At the same time I should like to thank all members of the $\alpha\beta$ sequencing group for being able to make use of the results prior to their publication. Among this group my special thanks go to Ms.A.Schmid for her expert technical assistance, and to Dr.A.Levanon for his numerous suggestions, scientific and otherwise.

My thanks are also extended to Dr.R.Portmann; without the computer programs and facilities he put at my disposal any sequence comparison and secondary structure calculation of RNA would have been impossible.

Finally my thanks go to all members of the Institute, past and present, for their helpfulness, their cooperation and their suggestions.

Contents

Summary

1. Introduction

- 1.1. General biology of bacteriophage $Q\beta$
- 1.2. Functional features of bacteriophage $Q\beta$ RNA
 - 1.2.1. Replication of genetic information
 - 1.2.2. Expression of genetic information
- 1.3. Template specificity of $Q\beta$ replicase
- 1.4. Hybrid plasmids containing $Q\beta$ DNA are a novel biological system
- 1.5. Nucleotide sequence determination of the bacteriophage $Q\beta$ genome

2. Materials and Methods

2.1. Materials

2.2. Methods

- 2.2.1. Preparation of double stranded DNA containing partial copies of the $Q\beta$ genome
 - 2.2.1.1. Partial cleavage of $Q\beta$ RNA with nuclease S1
 - 2.2.1.2. Elongation of RNA fragments with poly(A) tails
 - 2.2.1.3. Synthesis of the first $Q\beta$ DNA strand (minus strand)
 - 2.2.1.4. Synthesis of the second $Q\beta$ DNA strand (plus strand)
 - 2.2.1.5. Cleavage of the terminal hairpin loop of double stranded $Q\beta$ DNA with nuclease S1
 - 2.2.1.6. Elongation of the double stranded DNA with poly(dA)

- 2.2.2. Preparation of hybrid plasmids containing $Q\beta$ sequences
 - 2.2.2.1. Hybridization of the homopolymer-elongated DNA preparations
 - 2.2.2.2. Transfection of $Q\beta$ hybrid plasmid preparation into *E.coli*
- 2.2.3. Identification of bacterial clones harbouring plasmids with $Q\beta$ specific sequences
- 2.2.4. Isolation and characterization of $Q\beta$ plasmids
 - 2.2.4.1. growth of bacteria containing $Q\beta$ plasmids
 - 2.2.4.2. Isolation of plasmid DNA
 - 2.2.4.3. Excision of $Q\beta$ specific DNA inserts by nuclease S1
 - 2.2.4.4. Preparation of DNA restriction fragments
 - 2.2.4.5. Separation and isolation of DNA fragments by gel electrophoresis
- 2.2.5. Preparation of radioactively labelled DNA fragments for restriction mapping and nucleotide sequence determination
- 2.2.6. Determination of restriction sites in the $Q\beta$ DNA sequence
- 2.2.7. Determination of DNA nucleotide sequences

3. Results

- 3.1. Approach for the construction of hybrid plasmids containing partial copies of the $Q\beta$ genome
 - 3.1.1. generation of randomly cleaved $Q\beta$ RNA with 3'-poly(A) tails
 - 3.1.2. Synthesis of double stranded DNA containing sequences of $Q\beta$ RNA
 - 3.1.3. Elongation of double stranded $Q\beta$ DNA with poly(dA), annealing with poly(dT)-elongated plasmid PCRI and transfection of *E.coli*

3.1.4. Identification of bacteria harbouring hybrid plasmids with $\text{Q}\beta$ specific sequences and preliminary classification of selected plasmids

3.2. Restriction map of the $\text{Q}\beta$ genome

3.3. Nucleotide sequence determination of the $\text{Q}\beta$ genome

4. Discussion

4.1. Fidelity of $\text{Q}\beta$ RNA retrotranscription

4.2. Relationships between bacteriophages $\text{Q}\beta$ and MS2

4.2.1. Introduction

4.2.2. Comparison of coding sequences

4.2.2.1. Reading frame and codon usage

4.2.2.2. Comparison of coat-, replicase- and maturation proteins

4.2.3. Comparison of recognition sequences

4.2.3.1. Replicase binding sites

4.2.3.2. Coat protein binding sites

4.2.3.3. Nontranslated regions

4.2.4. Minor gene products of bacteriophages $\text{Q}\beta$ and MS2

4.2.4.1. Evolution of $\text{Q}\beta$ A1 protein

4.2.4.2. Comparison of a hypothetical lysis protein in phage $\text{Q}\beta$ with the actual lysis protein in phage MS2

4.2.4.3. Hybrid proteins arising by ribosomal frameshifting

References

Abbreviations

SUMMARY

The nucleotide sequence of 25% of bacteriophage $\varphi\beta$ RNA as a DNA transcript and the restriction map of the entire $\varphi\beta$ genome was determined, using PCRI- $\varphi\beta$ hybrid plasmids containing partial (Mekler et al., 1977) or complete (Taniguchi et al., 1978) copies of the $\varphi\beta$ genome. The whole sequence of 4220 nucleotides determined as a collaborative effort in our laboratory is presented and examined. A comparison of the coding sequences of bacteriophages $\varphi\beta$ and MS2 revealed that whereas maturation- and coat proteins show little to moderate aminoacid sequence homology, the replicase β -subunits of both phages show strong homology over 60% of the length of their aminoacid sequence. This sequence homology strongly supports the idea that phages $\varphi\beta$ and MS2 have evolved from a common ancestral phage, a notion which hitherto was based mainly on the similar layout of both phage genomes.

In the course of evolution however, each phage developed characteristic genetic features leading to the expression of specific minor proteins. In MS2, a "lysis protein" is translated in a second reading frame from an RNA segment forming part of two other genes; there is no evidence for a similar overlapping gene in $\varphi\beta$.

Possibly in $\varphi\beta$ this function relies on the A1 protein, which arises by occasional translation across the termination codon of the coat protein cistron into an adjacent sequence of 585 nucleotides.

The two phages show remotely similar primary and secondary structures in their terminal non-translated RNA regions. The $\varphi\beta$ RNA regions known to interact with $\varphi\beta$ replicase differ extensively from the MS2 RNA regions located in corresponding genome areas; the same was already known for the respective interaction sites of coat protein with the RNA.

Evidence for sequence duplication in the course of evolution is presented: in $\varphi\beta$ the A1 cistron clearly appears to have arisen from a large duplication; furthermore repeated short sequences in the extracistronic regions of both phages might indicate small duplication events.

1. INTRODUCTION

1.1. General Biology of Bacteriophage $\phi\beta$

Coliphages with an RNA genome may be classified into five groups, both by serological and biochemical criteria (Scott, 1965; Watanabe et al., 1967; Miyake et al., 1971).

The bacteriophages f2, R17 and MS2 are representatives of group I; they are quite similar among each other and differ considerably from group III phages, such as $\phi\beta$. The phages mentioned have been studied extensively, whereas phages of groups II, IV and V are not well characterized (Stavis & August, 1970; Weissmann et al., 1973; Zinder, 1975).

Phage particles are of icosahedral symmetry and are roughly spherical with a diameter of about 250 Å (Fischbach et al., 1965; Zipper et al., 1971). The particles consist of a protein shell encapsidating the single-stranded RNA genome. The RNAs are about 3500 (group I) to 4200 (group III) nucleotides long (Sinha et al., 1965; Boedtke, 1971) and have a high degree of secondary and tertiary structure (Boedtke, 1968; Fiers et al., 1976). The protein shell is made up of 180 coat protein molecules; additionally each capsid contains one molecule of maturation protein (A protein in group I phages, A2 protein in $\phi\beta$), which is essential for the phage binding to its host (Lodish et al., 1965; Roberts & Steitz, 1967; Krahn & Paranchych, 1971). Phage $\phi\beta$ contains a few molecules of yet another virus-specific protein, named A1 or IIa (Garwes et al., 1969). This A1 protein is necessary for the formation of a viable particle, as shown by reconstitution *in vitro* (Hofstetter et al., 1974).

By genetic analyses of group I and group III phages three regular cistrons have been identified: one coding for the maturation protein, a second for coat protein and a third one

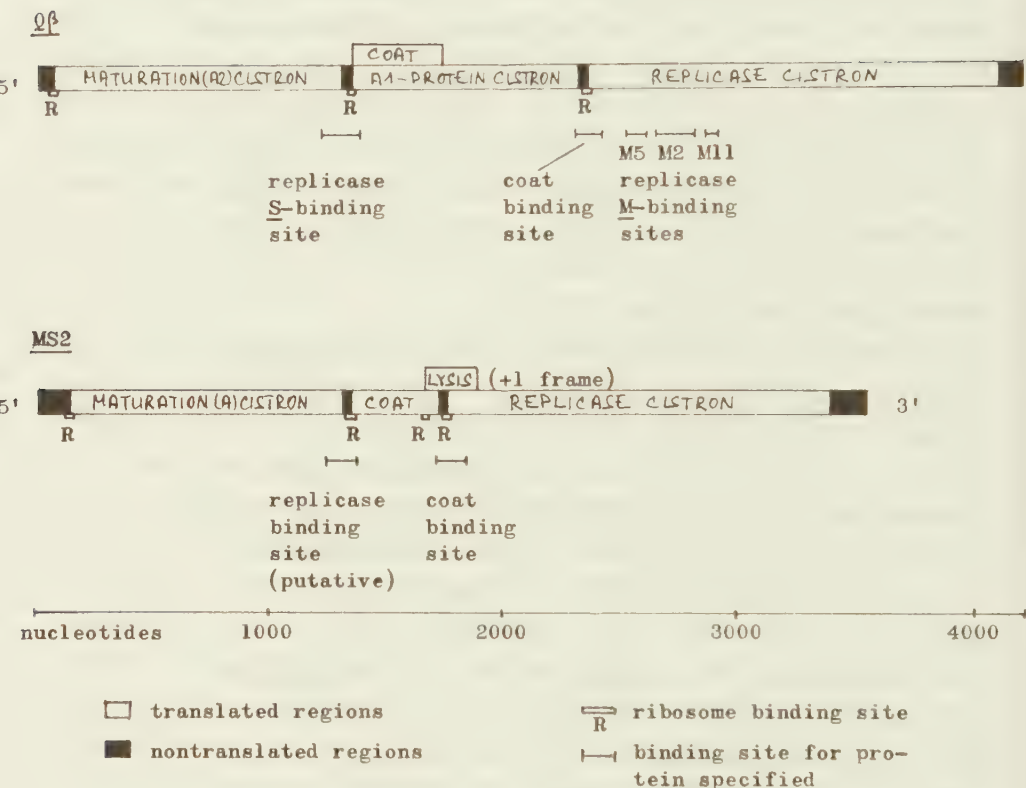
for the β -subunit of the viral RNA polymerase (Gussin, 1966; Horiuchi et al., 1966; Horiuchi & Matsubashi, 1970). Additional proteins are coded for by the phage genome in unusual ways (see fig.1). In $\phi\beta$ the A1 protein is formed when translation of the coat protein is not stopped at the termination codon, but continues into the adjacent region for about 600 nucleotides (Horiuchi et al., 1971). In group I phages a part the genome is read in a different frame, coding for a lysis protein (Atkins et al., 1979; Model et al., 1979), a situation found previously in single strand DNA phages (Barrell et al., 1976; Godson et al., 1978).

Several points are noteworthy:

(i) The RNA is subdivided into translated and nontranslated regions, as in the case of bacterial and eukaryotic genomes. At both 5'- and 3'-ends of the RNA there exist stretches of at least 60 nucleotides, and between any two cistrons segments of 20 or more nucleotides which are not translated into protein (Billeter et al., 1969a; Billeter et al., 1969b; Goodman et al., 1970; Hindley et al., 1970; Jeppesen et al., 1970; Nichols, 1970; Staples et al., 1971; Staples & Hindley, 1971). Regions immediately preceding the cistrons are required for initiation (ribosome binding sites) and regulation (replicase S-binding site, coat protein binding sites) of protein synthesis. The function of the long nontranslated segments at the 5'- and 3'-termini of the genomes is still unclear. However, point mutations generated in the 3'-terminal extracistronic region suggest that this stretch is essential for the biological activity of the phages. Alterations of one nucleotide can be lethal (Flavell et al., 1974; Sabo et al., 1977) or cause a selective disadvantage to the phage (Domingo et al., 1976).

(ii) Some nucleotide sequences are used for the synthesis of two different proteins: for $\phi\beta$ coat- and A1-polypeptides the

Figure 1: Genetic Map of Q β and MS2 Genomes



In the 4220 nucleotide Q β RNA, the cistron for maturation protein lies between positions 62 and 1321, respectively, that for the coat protein between 1345 and 1743, that for the A1-protein between 1345 and 2331 and that for the replicase β -subunit between 2353 and 4119; in the 3569 nucleotide MS2 RNA, the cistron for maturation protein lies between positions 130 and 1308, that for the coat protein between 1335 and 1724, that for the lysis protein between 1678 and 1902 and that for the replicase β -subunit between 1761 and 3395 (Fiers et al., 1976; Fiers, 1979).

same reading frame is exploited, whereas for MS2 coat-, replicase- and lysis proteins two different reading frames are used (see fig.1 and sections 4.3.1., 4.3.2.).

(iii) The similar organization of group I and III phage genomes and the hitherto noted slight homologies of some nucleotide sequences in certain regions suggested that the phages of both groups might be derived from a common ancestor (see sections 4.4., 4.3.1.).

The infectious cycle of RNA phages is subdivided into four different steps:

(i) In the adsorption step RNA phages attach only to the F-pili of "male" strains of E.coli (Loeb & Zinder, 1961; Crawford & Gesteland, 1964).

(ii) Penetration of phage RNA: after phage adsorption the RNA is released from the phage capsid by an unknown mechanism (Edgell & Ginoza, 1965) and migrates into the bacterium together with the maturation protein (Kozak & Nathans, 1971; Krahn et al., 1972). The RNA carries all the information required for successful infection: infection of bacterial spheroplasts with protein-free RNA gives rise to normal phage particles (Strauss, 1964).

(iii) In the maturation step the phage components are synthesized (see section 1.2.2.) and assembled spontaneously without requirement for extraneous morphopoietic factors. In vitro, infectious albeit physically abnormal phage particles can be reconstituted from isolated phage components (Roberts & Steitz, 1967; Hofstetter et al., 1974).

(iv) Lysis of the host bacterium occurs forty to sixty minutes after infection, releasing 10^4 to 4×10^4 particles, of which 10% to 50% are infectious. The mechanism of cell lysis is not well understood; in group I phages it is shown to depend on the function of the special lysis protein (Model et

al., 1979) mentioned above.

1.2. Functional Features of Bacteriophage $\phi\beta$ RNA

1.2.1. Replication of Genetic Information

Replication of the single stranded genomic RNA involves a two-step mechanism: the first step is the synthesis of a "minus" strand, viz. a strand complementary to the "plus" strand encapsidated in the phage (Weissmann & Ochoa, 1967; Weissmann et al., 1968a). This minus strand is then used in a second step as template for the synthesis of progeny phage RNA.

Both syntheses of plus and minus strand RNAs are carried out by the same phage-coded replicase (August et al., 1968; Weissmann et al., 1968b). Synthesis of the minus strand RNA starts at the 3'-end of the plus strand template (August et al., 1968; Spiegelman et al., 1968). In addition to replicase the E.coli-specific protein HFI is required to initiate synthesis of the minus strand (August et al., 1972; Franze de Fernandez et al., 1972). HFI has been shown to be a ribosome-associated protein with a high affinity for poly(rA) (Carmichael et al., 1975); its physiological function in E.coli is unknown. Synthesis of the plus strand starts likewise at the 3'-end of the minus strand RNA. For plus strand synthesis only replicase, but no HFI, is required.

The viral replicase (molecular weight 215'000) consists of four different subunits, only the β -subunit of which is coded for by the phage genome (Kamen, 1970; Kondo et al., 1970; Fedoroff & Zinder, 1971). The other three subunits (α -, γ -, δ -) are E.coli-specific polypeptides. The RNA-polymerizing activity resides in the phage-coded β -subunit (molecular weight 65'000 for $\phi\beta$, 63'000 for MS2; evidence for polymerase activity in Landers et al., 1974). The β -subunit is also

responsible for the high specificity of the replicase to its natural template (Haruna & Spiegelman, 1965a; Haruna & Spiegelman, 1965b; see also section 1.3.).

The largest subunit α (molecular weight 70'000) is the 30S ribosomal protein S1 (Kamen et al., 1972; Wahba et al., 1974) and is identical to the interference factor i (Groner et al., 1972). In the replicase holoenzyme the α -subunit is required for RNA synthesis directed by the phage plus strand RNA, but not by minus strands, 6S RNA or artificial templates. The S1-protein recognizes two sites in $\phi\phi$ plus strand RNA, one near the 3'-end and the other at the intercistronic region preceding the coat cistron (Senear & Steitz, 1976). The host factor HFI also interacts with $\phi\phi$ RNA near the 3'-end and near the end of the replicase cistron (Senear & Steitz, 1976). This suggests that protein S1 (and possibly HFI) is responsible for the binding of replicase to the 3'-terminal region of $\phi\phi$ RNA. The interaction of protein S1 with the intercistronic region preceding the coat cistron seems to be responsible for the role of $\phi\phi$ replicase as a translational repressor (Weber et al., 1972; see also section 1.2.2.).

The γ - and δ -subunits of $\phi\phi$ replicase are the elongation factors EF-Tu and EF-Ts of E.coli protein synthesis (molecular weight 45'000 and 35'000, respectively). They are responsible for correct initiation of RNA synthesis (Hori et al., 1974; Landers et al., 1974). In addition they seem to be required for RNA chain elongation (Carmichael et al., 1976). The exact roles of γ - and δ -subunits in $\phi\phi$ replication are unclear and seem totally unrelated to their normal function in E.coli protein synthesis.

1.2.2. Expression of Genetic Information

Upon entry into the bacterial cell the phage RNA first acts as a messenger for phage-specific protein synthesis (Godson & Sinsheimer, 1967). In vitro experiments with group I phages have shown that all phage cistrons are translated from the parental (plus) strand RNA (Lodish & Robertson, 1969; Kozak & Nathans, 1972; Sugiyama et al., 1972).

The various phage proteins are synthesized in different molar quantities, reflecting the differential need during phage replication and assembly. This coordinate control of translation is subject to four main regulatory mechanisms:

- (i) Static translational control due to RNA structure: Early in infection the secondary or tertiary structure of phage RNA causes initiation of protein synthesis to take place at a very different rate at each cistron. In native 30S $\phi\beta$ RNA ribosome binding and translation occur predominantly at the coat cistron. Initiation at the replicase cistron ensues only when the whole coat cistron is being actively translated (Gussin, 1966; Lodish & Zinder, 1966; Lodish, 1968). Here the translating ribosomes unfold the RNA and the ribosome binding site of the replicase cistron becomes accessible for initiation. Initiation at the maturation cistron takes place only on nascent plus strand RNA early during its synthesis. When the nascent plus strand becomes further elongated and folds into its secondary and tertiary structure the ribosome binding site of the maturation cistron is no longer accessible to ribosomes (Robertson & Lodish, 1970; Kolakofsky & Weissmann, 1971a; Kolakofsky & Weissmann, 1971b; Staples et al., 1971).
- (ii) Dynamic translational control due to RNA-protein interactions: Twenty to thirty minutes after infection replicase production is almost completely shut down (Viñuela et al., 1967). This is due to accumulation of coat protein, which binds to the region between coat and replicase cistrons

(Bernardi & Spahr, 1972). The bound coat protein molecules prevent further initiation at the replicase cistron (Robertson et al., 1968; Ward et al., 1968; Eggen & Nathans, 1969; Weber, 1976).

Another RNA-protein interaction plays a role in converting $Q\beta$ RNA from messenger into a template for replication. Early in infection $Q\beta$ RNA is covered with translating ribosomes; only after replicase has been produced by this process the RNA can be replicated. Since ribosomes during protein synthesis and replicase during RNA replication move in opposite directions over the RNA, and since replicase cannot dislodge ribosomes bound to $Q\beta$ RNA (Kolakofsky & Weissmann, 1971a), a special mechanism for stripping ribosomes off the RNA must exist. $Q\beta$ replicase displays a strong affinity for the so-called S-site region located at the ribosome entry site of the coat cistron (Weber et al., 1972). Replicase bound to the S-site prevents attachment of ribosomes to this site (Kolakofsky & Weissmann, 1971b). As mentioned above, ribosomes do not bind to the region preceding the maturation cistron in mature 30S $Q\beta$ RNA (Lodish & Robertson, 1969; Robertson & Lodish, 1970; Lodish, 1971; Staples et al., 1971), nor do ribosomes attach to the replicase cistron initiation site when translation of the coat cistron has ceased (Gussin, 1966; Lodish, 1968). Therefore all ribosome entry sites are blocked, the translating ribosomes run off the RNA and the RNA becomes available as a template for its own replication.

(iii) Leaky termination: As mentioned in section 1.1. the middle cistron of $Q\beta$ RNA codes for both coat and A1 protein. A1 protein arises by reading through the coat cistron termination signal into the adjacent nucleotide sequence at a low frequency (about 4%) (Horiuchi et al., 1971; Moore et al., 1971; Weiner & Weber, 1971; Weiner & Weber, 1973).

(iv) Translational modulation: The coat cistron is translated much more efficiently than the replicase or maturation cistrons, reflecting the predominant need for coat protein during phage development. Apart from the regulatory mechanisms discussed above it seems that the choice between homonymous codons for any given aminoacid is used to influence the rate of translation of a cistron. It has been pointed out by Fiers et al. (Fiers et al., 1976) that e.g. in the coat cistron of MS2 the codons AUU or AUC for ile are used exclusively, while the homonymous codon AUA is not found. In the maturation and replicase cistrons of MS2 the codon AUA for ile is used in 20% to 40% of all cases, together with AUU and AUC (see section 4.2.2.1.). It has been suggested that in cistrons to be translated efficiently codons (e.g. AUU, AUC) recognized by an abundantly available tRNA species are used, while in cistrons coding for proteins needed in low amounts a homonymous codon (e.g. AUA) recognized by a rare tRNA species is used. The proper choice of codons could therefore exert a "braking" or "accelerating" action on translation (Weissmann et al., 1973; Fiers et al., 1976).

1.3. Template Specificity of $\varphi\beta$ Replicase

An important property of $\varphi\beta$ replicase is its template specificity. Under in vitro conditions, simulating a "physiological" environment, only $\varphi\beta$ RNA (Haruna & Spiegelman, 1965a), its minus strand (Feix et al., 1968), $\varphi\beta$ "variant RNA" derivatives (Mills et al., 1967; Bishop & Spiegelman, 1968), "6S" RNA (Banerjee et al., 1969; Kacian et al., 1972; Prives & Silverman, 1972; Mills et al., 1973; Mills et al., 1975; Schaffner, 1977) or viral RNA of very closely related RNA phages (Miyake et al., 1971) are replicated by $\varphi\beta$ replicase. Other

RNAs, such as R17, f2 or MS2 phage RNAs, E.coli rRNA or tRNA are not replicated by $\Omega\beta$ replicase (Haruna & Spiegelman, 1965a; Haruna & Spiegelman, 1965b; Okada et al., 1971).

The molecular basis for the template specificity of $\Omega\beta$ replicase is illuminated by the following findings: 3'-terminal fragments of $\Omega\beta$ RNA are active as templates for replicase, provided that they are not shorter than 50% of the total RNA length (Schwyzer et al., 1972). Direct binding experiments of $\Omega\beta$ replicase to $\Omega\beta$ RNA revealed two internal binding sites on the RNA (Weber et al., 1972; Meyer, 1978). One site, the so-called S-site (Weber et al., 1972; see also section 4.2.3.1.) is apparently not essential for replication, since it is located at about 70% from the 3'-end of the total $\Omega\beta$ RNA length; this means the S-site is not present in fragments still active as templates. The other site, the M-region (Meyer, 1978; see also section 4.2.3.1.) is located around the middle of the $\Omega\beta$ genome and is essential for replication. Other RNAs, such as MS2 or f2 RNA, have relatively high affinities for $\Omega\beta$ replicase at an uncharacterized binding site (August et al., 1968; Silverman, 1973a; Silverman, 1973b; see also section 4.2.3.1.) and have 3'-terminal sequences identical to $\Omega\beta$ RNA (...CCCCA_{OH}). Yet these RNAs lack the discriminatory M-site and are therefore unable to be replicated by the $\Omega\beta$ enzyme. It should be noted that 6S RNA or some ribopolymers, containing C-residues either exclusively or at high proportions are also used as templates by $\Omega\beta$ replicase, but the protein-RNA recognition is probably very different in this case (Eikhom & Spiegelman, 1967; Hori et al., 1967; Feix & Hake, 1975; Küppers & Sumper, 1975).

1.4. Hybrid Plasmids Containing $\Omega\beta$ DNA Are a Novel Biological System

With the advent of DNA cloning techniques it was possible to transcribe the viral RNA into DNA and to integrate the DNA copies into carrier plasmids for propagation in *E. coli*. Such hybrid plasmids could be used for the determination of the nucleotide sequence of $\Omega\beta$ RNA: the nucleotide sequence determination of DNA by the new sequencing technique introduced by Maxam and Gilbert (Maxam & Gilbert, 1977; Maxam & Gilbert, 1979) is much faster and less laborious than RNA sequence determination (see section 1.5.). Even more importantly, such hybrid plasmids allow genetic manipulations which cannot be carried out on the RNA level and permit the preparation and study of unconditionally lethal $\Omega\beta$ mutant RNA species.

Different approaches for the preparation of $\Omega\beta$ hybrid plasmids have been chosen, viz. insertion of partial or complete copies of the $\Omega\beta$ genome into the carrier plasmid. In one approach, the RNA genome was cleaved at random into quarter-genome sized fragments and transcribed into double stranded DNA with reverse transcriptase (Mekler et al., 1977). This procedure was chosen to ensure the representation of the whole $\Omega\beta$ sequence among the library of $\Omega\beta$ hybrid plasmids, since it was anticipated that it would be impossible to transcribe the whole length of $\Omega\beta$ RNA with reverse transcriptase. The disadvantage of such an approach is twofold: first, the complete $\Omega\beta$ genome is represented as partial and overlapping copies in many different hybrid plasmids, and second, biological activity of such plasmids can hardly be expected in the bacterial host.

The other approach (Taniguchi et al., 1978) was to separately transcribe $\Omega\beta$ plus and minus strands into cDNA of close to full-length $\Omega\beta$ genome size with reverse transcriptase, and

to hybridize the minus and plus strand $\phi\phi$ DNA to yield incomplete double strands. Repair synthesis with DNA polymerase I then converted the terminal single-stranded regions into double strands; thus complete duplex DNA copies of the $\phi\phi$ genome were obtained. Insertion of this DNA into a carrier plasmid and transfection into *E.coli* surprisingly resulted in bacteria producing viable $\phi\phi$ bacteriophage (Taniguchi et al., 1978). Thus *E.coli* propagating an episomic $\phi\phi$ DNA and producing $\phi\phi$ phage constitutes a novel biological system. How is an infectious cycle initiated in a cell containing a $\phi\phi$ DNA hybrid plasmid? The first event must be the synthesis of a plus strand RNA transcript which then serves as a messenger for the synthesis of the phage-coded proteins, most importantly the replicase β -subunit (see sections 1.2.1. and 4.2.2.2.). It is unlikely that transcription by the DNA-dependent RNA polymerase of *E.coli* is initiated precisely at the beginning of the $\phi\phi$ sequence and is terminated at its end; among other possibilities it has therefore been proposed that an initially oversized transcript of the $\phi\phi$ plus strand (which probably could serve as a messenger) is processed in vivo to a size where it can be used as a template for $\phi\phi$ replicase; alternatively, the $\phi\phi$ enzyme initiates RNA synthesis on the oversized primary transcript at an internal site corresponding to the 3'-end of the phage genome (Taniguchi et al., 1978).

Once the first RNA transcript or replication product corresponding to the normal size $\phi\phi$ RNA has been made the usual infectious pathway of phage $\phi\phi$ is followed (see sections 1.1. and 1.2.).

It should be noted that with the proper experimental setup (site-directed mutagenesis (Müller et al., 1978), use of a conditionally active promotor to direct $\phi\phi$ RNA transcription,

purification and trimming of transcripts from the DNA regions containing mutagenized $\phi\beta$ sequences) the system permits in principle to study the effect of any base change on RNA function.

1.5. Nucleotide Sequence Determination of the Bacteriophage

$\phi\beta$ Genome

All the complex reactions and regulatory features mentioned so far made it desirable to determine the complete nucleotide sequence of phage $\phi\beta$ genome to understand the functions of this phage on the molecular level.

The advent of RNA sequence determination methods using uniformly [^{32}P]-labelled RNA (Sanger & Brownlee, 1967; Brownlee, 1972) brought forth a wealth of information. Starting with the elucidation of the ribosomal 5S RNA sequence of *E. coli* (120 nucleotides; Brownlee et al., 1967; Brownlee et al., 1968) its application culminated in the analysis of the entire MS2 RNA genome (3559 nucleotides; Nichols, 1970; De Wachter et al., 1971; Contreras et al., 1972; Min Jou et al., 1972; Rensing et al., 1973; Fiers et al., 1975; Fiers, 1975; Vandenberghe et al., 1975; Fiers et al., 1976).

In the case of $\phi\beta$ RNA almost 60% of its genomic nucleotide sequence was analyzed using this sequence determination method (Billeter et al., 1969a; Billeter et al., 1969b; Hindley & Staples, 1969; Goodman et al., 1970; Staples & Hindley, 1971; Steitz, 1972; Weber et al., 1972; Porter & Hindley, 1972; Kolakofsky et al., 1973; Porter et al., 1974; Weber, 1976; Billeter, 1978; Escarmis et al., 1978; Meyer, 1978; Shapira et al., 1978).

However, the analysis of long RNA sequences by the above methods is extremely laborious. On one hand, only fragments of 100 to 200 nucleotides can be analyzed at once. Since no specific methods of RNA cleavage at only few sites are available

the RNA under investigation is cleaved partially with ribonucleases and the very complex mixture of fragments has to be resolved by two or three subsequent gel electrophoresis separations. The isolated fragments are fully degraded to small oligonucleotides by ribonucleases of different cleavage specificities. The sequence of each oligonucleotide is then determined by a variety of degradation procedures. On the other hand, the order of these oligonucleotides within each fragment and finally the order of the fragments within the RNA has to be determined using sequence overlap data. This involves essentially the analysis of fragments ranging from a few up to thousands of nucleotides in length.

These difficulties are not encountered in a new method for nucleotide sequence determination of DNA (Maxam & Gilbert, 1977; Maxam & Gilbert, 1979). Unique fragments can be obtained using restriction enzymes. Furthermore it is not necessary to radioactively label the DNA *in vivo*, but DNA fragments of interest are labelled at the ends *in vitro*. The preparations are then subjected to a set of chemical modification and degradation reactions yielding many labelled fragments. After resolution by gel electrophoresis and autoradiography sequences of about 200 nucleotides from the labelled end can be read.

Therefore the method chosen to determine the nucleotide sequence of $\alpha\beta$ RNA was to first reverse transcribe the RNA into DNA (see section 1.4.) and then to determine the nucleotide sequence on the DNA level. The principle of this approach had proven successful for the sequence analysis of β -globin mRNA (Maniatis et al., 1976) and was therefore thought to be applicable to $\alpha\beta$ RNA as well.

2. MATERIALS AND METHODS

2.1. Materials

Enzymes: $\Omega\beta$ replicase, purified according to Kamen et al. (1972), was a gift of Prof.Dr.C.Weissmann; the specific activities used were 1900 U/mg (WR 323), 1400 U/mg (TW 335A) 1600 U/mg (TW 352). Host factor HFI was a gift of Prof.Dr.C.Weissmann; preparations used were TW 327B and TW 335A.

Terminal deoxyribonucleotide transferase (preparation Bo 272) and terminal ribonucleotide transferase were a gift of Prof.Dr.C.Weissmann. AMV reverse transcriptase was a gift of Dr.P.Curtis; T4-polynucleotide kinase was purchased from P&L Biochemicals.

Nuclease S1 was a gift of A.Schamböck. RNase A was purchased from Worthington; RNase T₁ (Calbiochem), treated with acid EDTA according to Dahlberg (1968), was a gift of Prof.Dr.M.A.Billeter.

Restriction endonuclease EcoRI was a gift of Prof.Dr.C.Weissmann; restriction endonuclease BspI was a gift of Dr.A.Kiss (Szeged). All other restriction enzymes were purchased from Biolabs or from P&L Biochemicals.

Pronase B grade (Calbiochem), incubated at 20 mg/ml in 5 mM EDTA, 10 mM Tris-HCl (pH 7.5) for 90 min at 37°C to destroy RNase activity, was a gift of Prof.Dr.M.A.Billeter. Lysozyme was from Calbiochem.

Nucleic acids: $\Omega\beta$ plus strand RNA (preparation V-84) was a gift of Prof.Dr.C.Weissmann. Yeast carrier RNA (British Drug Houses) was rendered free of RNase by extraction with phenol and precipitation with ethanol.

Globin cDNA was a gift of Dr.H.Hofstetter; poly(dT) elongated plasmid PCR1 was a gift of Dr.J.van den Berg. Sonicated calf thymus DNA was a gift of A.Schamböck.

[³²I]-labelled $\Omega\beta$ replicase binding $\Omega\beta$ RNA fragments M2 and S3 were a gift of Dr.F.Meyer.

Substrates for nucleic acid synthesis: Unlabelled ribonucleoside or deoxyribonucleoside triphosphate (sodium salt) were from Boehringer; γ - ^{32}P -dATP, α - ^{32}P -dGTP and ^3H -dATP were from NEN.

Material for chromatography and electrophoresis: Sephadex and Sepharose was from Pharmacia; poly(U) Sephadex G-10 was a gift of Prof. Dr. M. A. Billeter. Agarose (low EEO) was from Sigma. Cellulose DE-52 was from Whatman; hydroxyapatite (Hypatite) and Chelex-100 (sodium form) were from Biorad.

Filter papers (52, 3MM) were from Whatman; glass fibre filters (GF/C, 25 mm diameter) were from Whatman. Nitrocellulose filters (0.45 μm pore size, 25 mm diameter, type HAWP) and nitrocellulose sheets of the same specifications were purchased from Millipore.

Acrylamide (puriss.) and N,N'-methylene bisacrylamide (puriss.) were from Serva (Heidelberg, FRG).

Buffers: TNE (0.1 M Tris-HCl (pH 7.5), 0.15 M NaCl, 0.002 M EDTA) and SSC (0.15 M NaCl, 0.015 M sodium citrate (pH 6.45)) were autoclaved and stored at 4°C as ten- or twentyfold concentrated solutions, respectively (10xTNE, 20xSSC); they were diluted to the required final concentration with distilled water.

Tris (base), Tris-HCl and HEPES were from Sigma.

Material for bacteriological procedures: Normal medium (N-medium) contained Bacto-Tryptone (10 g; Difco), yeast extract (1 g; British Drug Houses), glucose (1 g), NaCl (8 g) and $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ (249 mg) per liter of solution; the medium was autoclaved at 120°C for 30 min to render it sterile. Normal agar (N-agar) was N-medium with 1.4% (w/v) of agar (granulate; BBL); the suspension of agar in the medium was autoclaved as described above, cooled down to 50-55°C (if required, kanamycin or streptomycin was added at this stage), and about 25 ml of this solution was used per Petri dish (plastic, 8.5 cm diameter). The agar plates

were cooled down to room temperature, dried upside down for 24 h at 37°C and stored at 4°C for several weeks.

Miscellaneous: BSA (fraction V, B grade; Calbiochem) was shaken with macaloid (Baroid National Lead Co., Houston) to remove RNase activity and dialyzed for several days against distilled water (gift of A.Schmid).

Phenol was distilled and stored at -20°C until use. Ether (Mallinckrodt, analytical reagent) was stored at -20°C; a few drops of water were added to each newly opened bottle to inactivate peroxydes. SDS was recrystallized from aqueous ethanol prior to use.

Bromophenol Blue was from Merck, Xylene Cyanol FF from British Drug House and ethidium bromide from Fluka.

X-ray films (RX PET) were from Fuji, Polaroid films (36 DIN) were from the Polaroid Company and photographic negative films (HP4, FP4) were from Ilford.

Tungstate radiation intensifier screens for autoradiography were from Ilford.

2.2. Methods

2.2.1. Preparation of Double Stranded DNA Containing Partial Copies of the $\phi\beta$ Genome

2.2.1.1. Partial Cleavage of $\phi\beta$ RNA with Nuclease S1

Full length [^{32}P] $\phi\beta$ plus strand RNA (210 μg of preparation V-84, from uncloned phage, containing 5×10^5 cpm of ^{32}P) was digested for 10 min at 37°C with nuclease S1 (0.003 U) in a total volume of 100 μl of 30 mM Na-acetate (pH 4.5), 50 mM NaCl, 1 mM ZnSO_4 and 5% (v/v) of glycerol. The reaction was stopped by addition of SDS (0.5% final concentration), and the mixture was further incubated for 10 min at 37°C with Pronase B (0.5 mg/ml; pretreated to remove nucleases, as specified in Methods). The mixture was extracted twice at room temperature with two volumes of phenol (equilibrated immediately prior to use with 1xTNE), and residual phenol was removed from the aqueous phase by three extractions with equal volumes of ice-cold ether. The RNA was precipitated from the aqueous solution by adding 0.1 volume of 10xTNE and 2.5 volumes of chilled absolute ethanol. After standing at -25°C for at least two hours the RNA was pelleted by centrifugation for 15 min at 10'000 rpm and 0°C in a Sorvall HB-4 swingout rotor. The RNA pellet was washed twice with 1 ml of 70% ethanol (ice-cold), centrifuged as above, and the ethanol was decanted carefully. Residual ethanol was dried away briefly in vacuo. The RNA pellet was dissolved in 1 mM EDTA to give a final concentration of 10 mg/ml, and the solution was stored frozen at -25°C . Recovery of the RNA was 98% of the input.

2.2.1.2. Elongation of RNA Fragments with Poly(A)-Tails

S1-digested RNA (200 μg , 1.5×10^5 cpm of ^{32}P) was dissolved in 500 μl of 0.2 M Tris-HCl (pH 8.3), 4 mM β -mercaptoethanol, 0.8 mM ATP, 0.5 mM MnCl_2 and 0.04 mg/ml BSA. Terminal thymus ribo-adenylate transferase (2.4 μg ; Gilvarg et al., 1975) was added and incubation was carried out for 180 min at 37°C . The reaction was stopped by addition of EDTA (20 mM final concentration), and the RNA was treated with SDS/pronase, phenol/ether extracted and precipitated as described in section 2.2.1.1.

After dissolving in 100 μl of 1 mM EDTA the RNA was denatured for 45 sec at 100°C and the solution was quenched in ice-water. NaCl, Tris-HCl (pH 7.5) and SDS (0.5 M, 10 mM and 0.5% final concentration, respectively) were added and the preparation was passed through a column of poly(U)-Sephadex G-10 (0.2 ml bed volume) and affinity chromatography was carried out as described by Coffin et al. (Coffin et al., 1974). The poly(A)-containing RNA eluting with 90% formamide was precipitated as described in section 2.2.1.1., the RNA was dissolved in 1 mM EDTA to give a final concentration of 1 mg/ml and the preparation was stored frozen at -25°C . The yield in this step was 26% of the input RNA (52 μg , containing 3.9×10^4 cpm of ^{32}P).

To eliminate small molecular weight RNA (<10S) the preparation (52 μg , 3000 Cherenkov cpm) was centrifuged through a 5-23% sucrose gradient (in 50 mM Tris-HCl, pH 7.5, and 1 mM EDTA) for 150 min at 55'000 rpm and 15°C in a Spinco SW60 rotor. The gradient was fractionated from the bottom, fractions corresponding to 12S to 20S RNA were pooled and the RNA was precipitated and washed as described in section 2.2.1.1. The RNA was dissolved in 0.1 mM EDTA to give a final concentration of 0.5 mg/ml . In this step 23 μg of poly(A) 23 RNA fragments were recovered (44% of input), containing 1320 Cher. cpm.

2.2.1.3. Synthesis of the First $\alpha\beta$ DNA Strand (Minus Strand)

For the synthesis of cDNA 5 μg of $\alpha\beta$ poly(A) RNA fragments and 1 μg of oligo(dT)₁₂₋₁₈ in 160 μl of 40 mM Tris-HCl (pH 8.0), 40 mM NaCl, 5 mM MgCl_2 , 0.5 mM dATP, 0.5 mM dCTP, 0.5 mM TTP, 0.3 mM α -[^{32}P]-dGTP (specific activity 2.5×10^5 cpm/nmol), 25 $\mu\text{g}/\text{ml}$ actinomycin D, 0.5 mM DTT, 0.3 mg/ml BSA were incubated with 20 U of AMV reverse transcriptase for 60 min at 42°C . The reaction was stopped by addition of EDTA (25 mM final concentration), and SDS/pronase/phenol treatment and precipitation of the nucleic acid was carried out as described in section 2.2.1.1. After dissolving in 100 μl of 20 mM Tris-HCl (pH 7.5), 1 mM EDTA the preparation was chromatographed through a column of Sephadex G-50 (2 ml bed volume, equilibrated with the sample buffer), fractions (100 μl each) were collected and identified by Cherenkov-counting, and the RNA in the preparation eluting in the exclusion volume of the column was digested with NaOH (0.3 M final concentration) for 16 h at 37°C . After neutralizing with 1 N HCl (0.27 volumes) the cDNA was precipitated by addition of two volumes of absolute ethanol and keeping the preparation at -25°C for 4 h. Centrifugation and washing of the DNA pellet was carried out as described for RNA preparations in section 2.2.1.1. In this step 1.2 μg of cDNA was obtained, containing 2.3×10^5 cpm of ^{32}P .

2.2.1.4. Synthesis of the Second $\alpha\beta$ DNA Strand (Plus Strand)

The single stranded cDNA (1.2 μg) was dissolved in 150 μl of 40 mM Tris-HCl (pH 8.0), 40 mM NaCl, 5 mM MgCl_2 , 0.5 mM dGTP, 0.5 mM dCTP, 0.5 mM TTP, 0.45 mM [^3H]-dATP (specific activity 1.6×10^5 cpm/nmol), 0.5 mM DTT, 0.3 mg/ml BSA and the mixture was incubated with 10 U of AMV reverse transcriptase for 120 min

at 42°C. The preparation was incubated with SDS/pronase and extracted with phenol as described in section 2.2.1.1., and chromatographed through Sephadex G-50 as described in section 2.2.1.3.

An aliquot (5%) of the pooled fractions from the Sephadex column was spotted onto a Whatman GF/C glass fibre filter and the ^{32}P - (minus strand DNA) and the ^3H -radioactivity (plus strand DNA) was determined by liquid scintillation counting. The DNA was precipitated with ethanol and washed as described in section 2.2.1.1.

In this step 0.5 μg plus strand DNA was obtained, containing 6×10^4 cpm of ^3H , and 1.1 μg minus strand DNA, containing 1.8×10^5 cpm of ^{32}P (91% overall recovery).

2.2.1.5. Cleavage of the Terminal Hairpin Loop of Double Stranded $\text{Q}\beta$ DNA with Nuclease S1

The bulk of the double stranded $\text{Q}\beta$ DNA preparation (0.4 μg of plus strand with 5×10^4 cpm of ^3H , 0.9 μg of minus strand with 9×10^4 cpm of ^{32}P) in 300 μl of 0.2 M NaCl, 50 mM Na-acetate (pH 4.5), 1 mM ZnCl_2 was incubated with 15 U of nuclease S1 (0.01 U of S1/ng DNA) for 30 min at 37°C. The reaction was stopped by adding EDTA (20 mM final concentration), the preparation was deproteinized, chromatographed through Sephadex G-50 and the DNA was precipitated as described in sections 2.2.1.1. and 2.2.1.3. 1.1 μg of DNA was obtained after this step, corresponding to 41'000 cpm of ^3H and 73'000 cpm of ^{32}P (recovery 81%).

The completeness of the S1-reaction was monitored by chromatography through hydroxyapatite. An aliquot from the excluded volume fractions of the Sephadex G-50 column (10% of total material) was brought to a final volume of 500 μl of 7.5 mM Tris-HCl (pH 7.5), 0.5% SDS, 40 $\mu\text{g}/\text{ml}$ yeast RNA, heat denatu-

red for 2 min at 100°C and quenched in ice-water. The preparation was adjusted to 1 ml of 0.14 M Na-phosphate (pH 6.8), 0.4% SDS, loaded onto a column of hydroxyapatite (0.2 ml bed volume, equilibrated in the sample buffer), and the DNA was eluted with 1 ml of 0.14 M Na-phosphate (pH 6.8), 0.4% SDS (single stranded DNA), and with 1 ml of 0.4 M Na-phosphate (pH 6.8), 0.4% SDS (double stranded DNA). Loading and elution of the hydroxyapatite column was carried out at 65°C. For analysis the DNA was precipitated on nitrocellulose discs (Millipore HAWP 45) with ice-cold TCA (10%), and the ^{32}P - and ^3H -radioactivity was determined by liquid scintillation counting.

90% of the ^3H - and 90% of the ^{32}P -radioactivity appeared in the single stranded fraction.

2.2.1.6. Elongation of the Double Stranded DNA with Poly(dA)

SI-treated double stranded λ DNA (80 ng, containing 2700 cpm of ^{32}P and 3000 cpm of ^3H) in 45 μl of 100mM K-cacodylate (pH 7.2), 4 mM MgCl_2 , 0.1 mM dATP and 50 $\mu\text{g}/\text{ml}$ BSA was incubated with 0.5 U of terminal deoxyribonucleotide transferase (preparation Bo 272) for 30 min at 37°C. The reaction was stopped with EDTA (20 mM final concentration) and the preparation was deproteinized as described in section 2.2.1.1.; 0.2 μg of ϕX174 DNA were then added as carrier. To remove unreacted dATP and small oligonucleotides the preparation was chromatographed through Sephadex G-50 as described in section 2.2.1.3. 44% (35 μg) of the input DNA was recovered. Half of this preparation was set aside for analysis on poly(U)-Sephadex (see section 3.1.3.); the other half, used for further preparative work, was precipitated and washed as described in section 2.2.1.1. The DNA was dissolved in 25 μl of 0.2xTNE and stored frozen at -25°C; the DNA amounted to 17 ng (700 cpm of ^3H , 600 cpm of ^{32}P).

2.2.2. Preparation of Hybrid Plasmids Containing $\Omega\beta$ Sequences

2.2.2.1. Hybridization of the Homopolymer-Elongated DNA Preparations

The kanamycin resistance plasmid PCRI (for review see Sinsheimer, 1976) of *E.coli* was cleaved with restriction endonuclease Eco RI at its unique site and tailed with poly(dT) at the 3'-ends (the cleaved and poly(dT) elongated plasmid was a gift from Dr.J.van den Berg). The annealing reactions were carried out in Eppendorf tubes, incubating 10 μg of the tailed plasmid preparation and various amounts of the poly(dA) elongated $\Omega\beta$ DNA (0.5 ng, 1 ng and 4 ng) in 25 μl of 1xTNE for initially 1 min at 60°C, then for 1 h each at 46°C, 37°C and at 20°C. The mixtures were adjusted to 60 mM Tris-HCl (pH 7.5), 10 mM MgCl_2 and 10 mM CaCl_2 (total volume was 30 μl), and the preparations were stored at 0°C until the recipient bacteria were ready for transfection (approximately 2 h).

No ligation procedure was required prior to transfection (Wensink et al., 1974).

2.2.2.2. Transfection of $\Omega\beta$ Hybrid Plasmid Preparation into *E.coli*

Transfection was carried out according to the method of Mandel & Higa (Mandel & Higa, 1970). A freshly inoculated culture of *E.coli* N543 (50 ml culture volume, inoculum 1:25) in normal medium (see Materials) containing streptomycin (200 $\mu\text{g}/\text{ml}$) was incubated at 37°C until the density had reached approximately 2×10^8 cells/ml (about 2 h). The bacteria were spun down at 0°C and 3000 rpm for 5 min in the Sorvall HB-4 rotor, the supernatant was decanted and the inside of the tubes was carefully dried with a sterile paper tissue. The bacterial pellets were gently resuspended by manual agitation in

10 ml of sterile CaCl_2 solution (50 mM) and kept on ice. Immediately afterwards the $\text{Q}\beta$ DNA-plasmid hybrid preparation (see section 2.2.2.1.) was transferred into a sterile 3 ml plastic tube, and the remainder of the DNA preparation was transferred to this mixture with another 20 μl of 60 mM Tris-HCl (pH 7.5), 50 mM NaCl, 10 mM MgCl_2 , 10 mM CaCl_2 . To this DNA preparation 50 μl of the CaCl_2 -treated recipient bacteria was added, the mixture was incubated for 1 h at 0°C , then 1 min at 42°C , and after cooling to 0°C normal medium containing streptomycin (1 ml of medium, 200 $\mu\text{g}/\text{ml}$ streptomycin) was added. The incubation was continued at 37°C for 90 min. Two portions of this mixture (0.5 ml each) were plated on normal agar containing kanamycin (50 $\mu\text{g}/\text{ml}$) and streptomycin (200 $\mu\text{g}/\text{ml}$), and the petri dishes were incubated at 37°C for 18 h, when bacterial colonies could clearly be observed and counted.

2.2.3. Identification of Bacterial Clones Harbouring Plasmids with $\text{Q}\beta$ Specific Sequences

To identify bacterial clones containing plasmids with $\text{Q}\beta$ specific sequences a simplified Grunstein-Hogness procedure was used (Grunstein & Hogness, 1975).

Discs cut from nitrocellulose sheets (Millipore HA, 45 μm pore size, 82 mm diameter) were washed three times in boiling water, dried and autoclaved for 10 min at 120°C . Colonies were gridded onto the filter disc using sterile wooden toothpicks, the filter was placed on a petri dish containing normal agar with kanamycin (50 $\mu\text{g}/\text{ml}$) and streptomycin (200 $\mu\text{g}/\text{ml}$), and incubation was carried out at 20°C until the colonies reached a diameter of approximately 2 mm (60 to 70 h). The filter was placed on top of two layers of Whatman 3MM paper soaked with 0.5 M NaOH for 10 min at 20°C and then transfer-

red to an identical arrangement of 3MM paper soaked with 1.0 M Tris-HCl (pH 7.5). Every 10 min such transfers were repeated until the solution in contact with the nitrocellulose disc had reached neutrality (about three transfers). Lastly, the disc was left in contact with 1.5 M NaCl, 0.5 M Tris-HCl (pH 7.5) for 5 min, and the disc was dried on a vacuum funnel. The disc was then overlaid with a solution of pronase B (0.5 mg/ml) for 15 min at 20°C, the enzyme solution was sucked off and 96% ethanol (1 ml/cm² of disc surface) was sucked through. After five similar washing cycles with chloroform (2 ml per cm² of filter disc surface) the filter was dipped into 0.3 M NaCl to remove loose debris. The DNA was baked into the nitrocellulose discs at 80°C for 2 h in vacuo (12 mm Hg). At this stage the filters could be stored in vacuo at room temperature for several months.

Hybridization with ³²P-labelled $\alpha\beta$ specific probes (specific activity 10⁵-10⁶ cpm/pmol) was carried out as follows: 800 μ l of 50% formamide in 5xSSC buffer, containing 10⁶ to 10⁷ cpm of the radioactive probe and 5 μ g of yeast RNA, were used to wet the nitrocellulose filter with the baked-in DNA, and incubation was carried out at 37°C for 16 h in a plastic petri dish (85 mm diameter) under a layer of paraffine oil (for oligonucleotide probes of 15 or less nucleotides length the incubation was for 16 h at 37°C, then for 4 h at 30°C and 4 h at 25°C; see Szoslak et al., 1979). The filter was washed free of paraffine oil with three portions of chloroform (100 ml/filter) for 10 min at room temperature under slight manual agitation; finally the filters were washed successively with 6xSSC, 2xSSC, 2xSSC containing RNase A (5 mg/ml) for 10 min at 20°C (100 ml of solution/wash/filter).

The filter was blotted, air-dried autoradiographed at -70°C using a tungstate intensifier screen.

2.2.4. Isolation and Characterization of $\phi\beta$ Plasmids

2.2.4.1. Growth of Bacteria Containing $\phi\beta$ Plasmids

Clones showing hybridization with $\phi\beta$ -specific probes (see section 2.2.3.) were picked off the original petri dish, grown in 0.5 ml of normal medium containing kanamycin (50 μg per ml) and streptomycin (200 $\mu\text{g}/\text{ml}$) up to a density of approximately 10^8 cells/ml, when sterile glycerol was added (30% final concentration). These glycerol cultures could be stored at -25°C for several months.

The growth protocol for preparative plasmid isolation was as follows: 1 l of normal medium containing kanamycin (50 $\mu\text{g}/\text{ml}$) and streptomycin (200 $\mu\text{g}/\text{ml}$) was inoculated with 20 ml to 25 ml of a starter culture (approximately 3×10^8 cells/ml) and the 1 l-culture was agitated at 37°C (shaking platform at 80-100 rpm) until the cell density had reached approximately 3×10^8 cells/ml (4-7 h of incubation). Chloramphenicol (a concentrated stock solution in 96% ethanol) was added to a final concentration of 0.17 mg/ml, and incubation was continued at 37°C for another 18 h to amplify the number of plasmid copies per cell (Chang & Cohen, 1978). The bacteria were then centrifuged for 15 min at 5000 rpm and 4°C in plastic flasks (300 ml) in a Sorvall GSA rotor. The bacterial pellet (3-5 g wet cell weight per liter of culture) was washed twice with 20 mM Tris-HCl (pH 8.0; 10 ml of buffer per gram of wet cell weight), centrifuged for 10 min at 8000 rpm and 4°C , and the pellet was used immediately for isolation of plasmid DNA (see section 2.2.4.2.) or stored in 15 ml of 25% sucrose, 20 mM Tris-HCl (pH 8.0) at -25°C until used for DNA isolation.

2.2.4.2. Isolation of Plasmid DNA

Three different procedures for the isolation of plasmid DNA were used, differing mainly in the method of bacterial lysis and the work-up of the DNA.

(i) For lysozyme-SDS lysis the bacterial pellet was suspended in 50 mM Tris-HCl (pH 8.0; 2 ml of buffer per gram of wet cell weight), lysozyme was added to a final concentration of 1.2 mg/ml and the suspension was kept at 0°C until it had become slightly viscous (approximately 5 min). EDTA was added (25 mM final concentration), and after 5 min at 0°C the preparation was adjusted to 0.85 M NaCl, 0.7% SDS (w/v). After 15-18 h at 4°C the high molecular weight DNA and cellular debris was pelleted for 60 min in a Spinco SW-27 rotor at 25'000 rpm and 4°C, the supernatant containing RNA and low molecular weight DNA was carefully removed and thoroughly shaken with 0.5 volume of phenol for 5 min at room temperature. Chloroform (0.2 volume) was added, the emulsion agitated for 5 min and the phases were separated by centrifugation at 6000 rpm and 15°C in a Sorvall HB-4 rotor. The phenol-chloroform extraction step was repeated twice, and the nucleic acid was precipitated from the aqueous phase with ethanol as described in section 2.2.1.1.

(ii) In the alternative lysozyme-Triton method the bacterial pellet was suspended in 50 mM Tris-HCl (pH 8.0; 10 ml of buffer per gram of wet cell weight), lysozyme was added (2 mg/ml final concentration), after 10 min at 0°C the now slightly viscous slurry was adjusted to 100 mM EDTA, after 10 min at 0°C Triton X-100 was added (0.11% final concentration, v/v) and the preparation was kept at 0°C for 60 min. After centrifugation in a Sorvall HB-4 rotor for 60 min at 0°C and 11'000 rpm the clear supernatant was transferred to a glass-beaker and the DNA was subjected to alkaline denaturation/renaturation (Currier & Nester, 1976; outlined in Wilkie et al., 1979):

there the pH of the solution was adjusted to pH 12.45 with 5 M NaOH and stirred slowly for 10 min at room temperature. The pH was lowered to pH 8.5 with 2 M Tris-HCl (pH 7.3), NaCl was added (0.5 M final concentration) and the solution was extracted with 2 volumes of phenol (equilibrated immediately prior to use with 0.5 M NaCl, 50 mM Tris-HCl, pH 8.5). After centrifugation in a Sorvall HB-4 rotor for 10 min at 6000 rpm and 4°C the aqueous upper phase was removed, carefully the granular interphase. The aqueous phase was further extracted with 1 volume of chloroform, centrifuged briefly for phase separation and the nucleic acid in the aqueous phase was precipitated with ethanol as described in section 2.2.1.1.

In both lysis methods ((i) and (ii)) RNA was removed from the preparation by dissolving the nucleic acid pellet in 1xTNE (2 ml of buffer per mg of nucleic acid), RNase A was added (10 µg/ml final concentration; contaminating DNase in the RNase preparation had been inactivated by incubation in 1xTNE at 90°C for 10 min), and the preparation was incubated for 60 min at 37°C. NaCl (0.4 M final concentration) and PEG-6000 (10% final concentration) was added, the preparation was kept at 0°C for 2 h, centrifuged for 10 min at 10'000 rpm and 0°C in a Sorvall HB-4 rotor and the pellet was dissolved in 50 mM Tris-HCl (pH 7.5), 1 mM EDTA (2 ml of buffer per mg of nucleic acid). RNase was removed by pronase/SDS/phenol treatment and the DNA was precipitated with TNE and ethanol as described in section 2.2.1.1. In method (i) the DNA was at this step subjected to the alkaline denaturation/renaturation and phenol treatment described above for method (ii).

To separate form I from form II and III plasmid DNA the preparation was dissolved in 50 mM Tris-HCl (pH 7.5), 1 mM EDTA

(to give a DNA concentration of 0.3-0.5 mg/ml) and centrifuged through a 5-23% sucrose gradient (in 50 mM Tris-HCl, 1 mM EDTA (pH 7.5)) for 18 h at 21'000 rpm and 15°C in a Spinco SW-27 rotor. The gradient was fractionated from top to bottom, absorption at 254 nm was determined in an ISCO gradient fractionator (1.2 ml/fraction, 3 ml/min flow rate, typically 30 to 32 fractions), the fractions corresponding to form I plasmid DNA were pooled and the DNA was precipitated with TNE and ethanol as described in section 2.2.1.1. Yields in plasmid isolation ranged between 20-200 µg of form I plasmid DNA per liter of starting culture.

(iii) When only small amounts of plasmid DNA from many different bacterial clones were required, the plasmid isolation protocol described in (ii) was scaled down to smaller starter cultures (30 ml), using the lysozyme-Triton lysis method, but omitting the alkaline denaturation/renaturation step of the DNA and the sucrose gradient centrifugation. Yields found in this small-scale method were 0.2-2 µg of form I plasmid DNA.

2.2.4.3. Excision of $\phi\beta$ Specific DNA Inserts by Nuclease S1

The $\phi\beta$ DNA flanked by the poly(dA).poly(dT) segments was excised from the hybrid plasmids according to the protocol of Hofstetter et al. (Hofstetter et al., 1976).

For preparative work 50 µg of plasmid DNA in 500 µl of formamide (40-60% w/v; the particular formamide concentration had been determined analytically in a preliminary experiment) containing 30 mM Na-acetate (pH 4.0), 200 mM NaCl, 1 mM ZnSO₄ were digested with 150 U of nuclease S1 for 30 min at 50°C. The mixture was adjusted to 5 mM EDTA, the DNA preparation was extracted three times with phenol (equilibrated immediately prior to use with 100 mM Tris-HCl (pH 7.5), 200 mM NaCl) and precipitated at -25°C by adding two volumes of absolute ethanol.

For the separation of the excised DNA sequence from the large parental plasmid DNA the preparation was dissolved in 100 μ l of 50 mM Tris-HCl (pH 7.5), 1 mM EDTA and centrifuged through a 5-23% sucrose gradient in a Spinco SW-60 rotor for 210 min at 54'000 rpm and 15°C. The DNA in the fractions of the sucrose gradient was identified by electrophoresis of an aliquot in an analytical agarose gel (see section 2.2.4.5.), the fractions containing the DNA fragment of interest were pooled and the DNA was precipitated with TNE and ethanol as described in section 2.2.1.1. Recovery of the $\Omega\beta$ DNA with this method was 70-85% of the theoretical molar yield.

2.2.4.4. Preparation of DNA Restriction Fragments

For digestion with restriction enzymes a series of standard buffers ("Y-buffers", "Z-10 buffer") was prepared; they contained 10 mM Tris-HCl (pH 7.5), 6 mM $MgCl_2$, 6 mM β -mercaptoethanol, 200 μ g/ml BSA (or gelatin), and varying amounts of NaCl: no NaCl (Y-0), 10 mM NaCl (Y-10), 50 mM NaCl (Y-50), 100 mM NaCl (Y-100) or 150 mM NaCl (Y-150). Z-10 buffer contained 10 mM Tris-HCl (pH 8.5), 6 mM $MgCl_2$, 6 mM β -mercaptoethanol, 200 μ g/ml BSA (or gelatin) and 10 mM NaCl.

Buffers used for the various restriction enzymes were Y-0 (AvaI, AvaII), Y-10 (BclI, BglII, CeuI, CeuII, HaeII, HapII, HpaII, KpnI, MboII, PvuII, RspI, TaqI), Y-50 (AluI, BspRI, HindIII, HinfI, PstI), Y-100 (BamHI, EcoRI, SalI), Y-150 (HaeIII, HhaI, MboII, PvuI, XhoI, XbaI) and Z-10 (SmaI). For nomenclature and sequence specificities of these enzymes see the review of Roberts (Roberts, 1979).

Incubation of the DNA with the restriction enzyme of interest was carried out as follows: up to 2 μ g of DNA per 10 μ l of reaction volume was mixed with 0.2 reaction volume of 5xY- or 5xZ-buffer and H_2O , the appropriate amount of restriction

enzyme was added and the incubation was carried out at 37°C (65°C for TaqI). The amount of restriction enzyme and the time of incubation had been determined in preliminary experiments. For preparative work the incubation was terminated by lowering the temperature to 0°C and adding EDTA (20 mM final concentration), and SDS/pronase/phenol treatment and precipitation of the DNA was carried out as described in section 2.2.1.1. For analytical work the incubation was terminated by adding 0.2 reaction volume of a solution containing 70% sucrose, 100 mM EDTA, 2% SDS and bromophenol blue (0.1 mg/ml); such preparations were used for analytical electrophoretic separations without further work-up (see section 2.2.4.5.).

2.2.4.5. Separation and Isolation of DNA Fragments by Gel Electrophoresis

Two types of gels were used for separation by electrophoresis, viz. agarose and polyacrylamide gels (for a review of gel electrophoretic separation methods see Southern (Southern, 1979)).

Electrophoresis in agarose gels was carried out in horizontal or in vertical gel systems. Horizontal gels (9.5x8x0.8 cm, top side uncovered and exposed to ambient air) were mainly used for quick analytical separations. Vertical gels (20x20x0.2 cm) immersed in the buffer compartment were used in preparative work. In both systems the buffer used was 50 mM Tris-acetate (pH 7.8), 2 mM EDTA, and the field applied was 5-10 V/cm. Agarose concentrations ranged from 0.5% to 2% (w/v) and were optimized for each particular separation problem (0.5% gels gave an optimal separation for fragments ranging from 1500 to 5000 base pairs, 1% gels for fragments of 700-1500 base pairs and 2% gels for fragments of 150-700 base pairs).

Electrophoresis in polyacrylamide gels was used for analytical and preparative work involving the separation of complex mixtures of DNA fragments in the range of 30 to 1500 base pairs. Standard gel dimensions were 20x40x0.2 cm, and the buffers used were either 50 mM Tris-borate (pH 8.3), 1 mM EDTA (Peacock & Dingman, 1967) for gels above an acrylamide concentration of 10%, or 36 mM Tris-phosphate (pH 7.8), 1 mM EDTA in all other cases (Loening, 1967).

Acrylamide concentrations used were 3-12% (acrylamide:bisacrylamide 30:1), and the concentrations had to be optimized for each particular separation problem (5% gels were optimal for the separation of fragments of 70-500 base pairs, whereas 10% gels were appropriate for the separation of fragments of 30-300 base pairs). The catalysts used in gel polymerization were ammoniumperoxysulphate (0.05%) and TEMED (0.02%). Electrophoresis was carried out in a vertical position (gel immersed in the buffer compartment for efficient cooling) at room temperature and at a field of 5-10 V/cm.

Radioactive DNA fragments were localized in the gels by autoradiography. Gels containing unlabelled fragments were stained for 3-10 min in a solution of running buffer containing ethidium bromide (0.4 ug/ml). The DNA was visualized under UV light (252 nm wavelength) and the gel was photographed through a Kodak Wratten filter, using Polaroid 36 DIN film (exposure time 0.5-5 sec).

In preparative work, the isolation and purification of DNA fragments was carried out as follows: the band containing the DNA fragment was cut out from the (agarose or polyacrylamide) gel, the gel piece was crushed by squeezing it through a syringe (2 ml, one-way) without a needle, and the DNA was extracted from the crushed gel with 10 gel volumes of 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5 mM EDTA by sha-

king the preparation, in the presence of yeast carrier RNA (5-10 μg), at room temperature for 10-16 h. After a brief centrifugation in the table-top centrifuge the supernatant was collected and the gel pellet was reextracted with five gel volumes of the extraction buffer described above for 2-4 h. The collected supernatants were filtered through a glass wool plug to remove residual gel particles, and the DNA was adsorbed to a column of Whatman DE-52 (containing a thin bottom layer of Chelex-100; total bed volume was 0.2 ml). The column was washed (10 bed volumes of 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5 mM EDTA) and the DNA was eluted from the column with 0.2-0.4 ml of 50 mM Tris-HCl (pH 7.5), 1.5 M NaCl, 5 mM EDTA. The solution was diluted with H_2O to give a final concentration of 0.75 M NaCl, and the DNA was precipitated by addition of 2 volumes of absolute ethanol at -25°C .

In some instances (when the DNA was to be treated further with a restriction enzyme (e.g. HhaI) sensitive to impurities eluted from the gel and not eliminated in this procedure) a different purification scheme was adopted. The DNA preparation, after elution from the gel as described above, was adsorbed to a hydroxyapatite column at room temperature (0.2 ml bed volume; the column had been equilibrated with 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5 mM EDTA prior to use). The column was washed with 10 bed volumes of 0.1 M phosphate buffer (pH 7.5), and the DNA was eluted from the column dropwise with maximally 2 bed volumes of 1 M phosphate buffer (pH 7.5). The preparation was diluted with H_2O (0.1 M phosphate final concentration), applied directly to a DE-52 column and eluted as described above.

Overall yields after gel extraction and DE-52 columns were 60-80%. Where purification required initial passage through hydroxyapatite overall yields dropped to 50-65%.

2.2.5. Preparation of Radioactively Labelled DNA Fragments
For Restriction Mapping and Nucleotide Sequence De-
termination

Restricted DNA (usually about 5-10 μg) was 5'-terminally labelled with [^{32}P]-phosphate as described by Mantei (Mantei et al., 1980).

To produce DNAs which carried [^{32}P]-phosphate groups at only one 5'-terminus two approaches were chosen. One approach consisted in cleaving the restriction fragment labelled at both 5'-ends with a suitable restriction enzyme, followed by separation and isolation of the unequally sized products by gel electrophoresis as described in section 2.2.4.5.

The other method entailed denaturation of the double stranded DNA, annealing to an excess of unlabelled $\varrho\beta$ plus strand RNA and separation of the RNA-DNA hybrid from the single DNA plus strands by gel chromatography.

The DNA preparation (2-5 μg , about 5 pmol) was dissolved in 300 μl of 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, full-length $\varrho\beta$ plus strand RNA (500 μg , about 50-70-fold molar excess over the DNA) and 1 mg of yeast carrier RNA were added, and the mixture was held at 100°C for 2 min for heat denaturation (nuclease-free EDTA treated pyrex tube). After quenching at 0°C the mixture was adjusted to 70% formamide, 0.5 M NaCl, 50 mM HEPES (pH 7.5), 10 mM EDTA in a total volume of 2.5 ml and incubated at 37°C for 60 min to allow RNA-DNA hybridization to take place. The preparation was diluted to 5 ml with H_2O , and the nucleic acid was precipitated at -25°C in the presence of two volumes of absolute ethanol. The nucleic acid was centrifuged and washed as described in section 2.2.1.1., and the pellet was dissolved in 300 μl of 1xTNE. The solution was applied to a column of Sepharose CL-4B or CL-6B (10 ml bed volume; the Sepharose had been autoclaved

and equilibrated with 1xTNE prior to use. For DNA fragments below 300 base pairs length Sepharose CL-6B was used, while fragments in the range of 300 to 800 base pairs necessitated the use of Sepharose CL-4B for optimal separation). The column was run at room temperature, a flow rate of about 5 ml/h was maintained and fractions of 300 μ l were collected. Radioactive DNA eluting in the excluded volume (RNA-minus strand DNA hybrid) or in the included volume (single stranded plus strand DNA) was monitored by measuring the Cherenkov radiation of the fractions, the fractions of interest were pooled and the nucleic acid was precipitated at -25°C with two volumes of absolute ethanol (recovery of the DNA was 65-80% of the input). The preparation was centrifuged and washed as described in section 2.2.1.1., the pellet was dissolved in 200 μ l of 0.3 M NaOH, and the RNA was hydrolyzed at 37°C for 4 h. The pH of the preparation was lowered to pH 7.0 with 0.6 M acetic acid, sonicated calf thymus carrier DNA (10 μ g) was added and the DNA was precipitated at -25°C with 2 volumes of absolute ethanol. Crosscontamination of the two complementary DNA strands was less than 0.1%.

2.2.6. Determination of Restriction Sites in the $\text{Q}\beta$ DNA Sequence

Two methods were used to determine the restriction map of $\text{Q}\beta$ DNA. The first method entailed complete cleavage of unlabelled DNA with a combination of different restriction enzymes. Double or triple cleavage with BamHI, BglII, BstEI, EcoRI, PvuII, RspI, SstI and XhoI enzymes as well as separation and length measurements of the products by gel electrophoresis was carried out as described in the figure legends and in sections 2.2.4.4. and 2.2.4.5. DNA length standards for coelectrophoresis were fragments

arising from the digestion of plasmid pBR 322 with restriction endonuclease HaeIII; the exact length of the DNA fragments was known by sequence determination (Sutcliffe, 1978a; Sutcliffe, 1978b).

Alternatively, Smith-Birnstiel restriction mapping (Smith & Birnstiel, 1976) was carried out. Fragments labelled with [32 P]-phosphate at the three BamHI sites of ϕ DNA and cleaved with restriction enzyme XhoI were used for this purpose. To partially cleave the DNA fragments labelled at one 5'-end the amount of restriction enzyme required for complete digestion was reduced 10-100-fold, and the digestion products were sampled at various time points (0-240 min). For unstable enzymes (e.g. HhaI) the ratio of enzyme to DNA rather than the incubation time was varied. In the case of particularly active restriction enzymes (e.g. TaqI) both the ratio of enzyme to DNA and the temperature of incubation were lowered to suboptimal values (for TaqI from 65°C to 30°C) to obtain all possible partial digestion products. Separation and length measurements of the products was carried out as described in the figure legends, in section 2.2.4.4. and in section 2.2.4.5.

2.2.7. Determination of DNA Nucleotide Sequences

The DNA fragments labelled at one 5'-end with [32 P]-phosphate were chemically degraded using the base-specific reactions according to Maxam and Gilbert (Maxam & Gilbert, 1977; reviewed in Maxam & Gilbert, 1979). The products after chemical degradation were fractionated on polyacrylamide gels (40x28x0.1 cm) in 50 mM Tris-borate (pH 8.3), 1 mM EDTA, 7 M urea. Acrylamide concentrations used were 20% for separation of products up to 150 nucleotides in length (acrylamide:bisacrylamide 30:1) and 12% for products longer than 150 nucleotides

(acrylamide:bisacrylamide 18:1). Runs were for 3, 12 and 25 h at 900 V, following a 12 h prerun at 700 V. For autoradiography the gels were transferred to aluminium plates (40x20x0.3 cm), covered with Saran-wrap and the X-ray film was exposed for 2-14 days at -70°C , using tungstate intensifier screens.

3. RESULTS

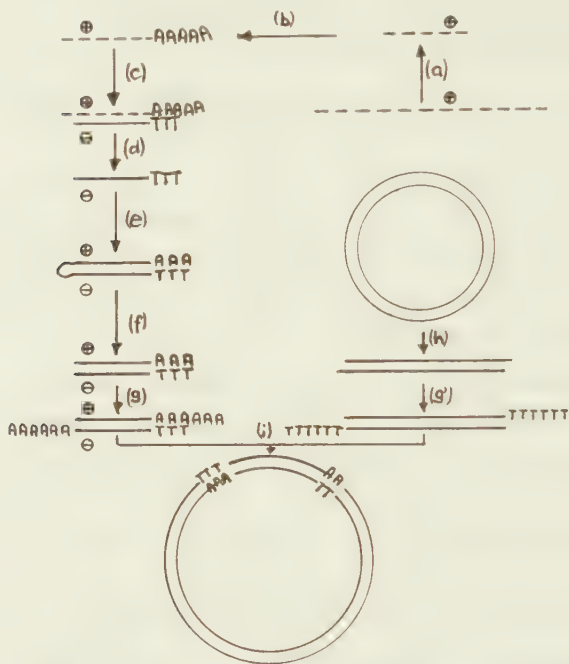
3.1. Approach for the Construction of Hybrid Plasmids Containing Partial Copies of the $\Omega\beta$ Genome

As mentioned in section 1.4. it was anticipated that it would be impossible to transcribe the whole length of $\Omega\beta$ RNA with reverse transcriptase. To nevertheless ensure the representation of the whole $\Omega\beta$ sequence in the DNA transcripts the approach outlined in fig.2 was adopted. Full size $\Omega\beta$ plus strands were partially cleaved with nuclease S1 to yield random fragments of approximately 1000 nucleotides. These fragments were then elongated with Ap residues using terminal riboadenylate transferase (Gilvarg et al., 1975) and used as templates for reverse transcriptase in an oligo(dT) primed reaction. The cDNA copies were freed of RNA by alkaline digestion and the DNA strand complementary to the cDNA was synthesized with reverse transcriptase in a second, unprimed reaction. The resulting hairpin covalently linking the two DNA strands was cleaved with low amounts of nuclease S1, the double stranded $\Omega\beta$ DNA was elongated with dA residues, hybridized to poly(dT)-PCR1 DNA and transfected into E.coli N543. 240 kanamycin-resistant colonies were gridded and assayed by the Grunstein-Hogness colony hybridization assay (Grunstein & Hogness, 1975), using [^{32}P]- $\Omega\beta$ RNA as a probe.

3.1.1. Generation of Randomly Cleaved $\Omega\beta$ RNA with 3'-Poly(A)-Tails

Conditions in the cleavage reaction were chosen such as to yield RNA fragments of approximately 1000 nucleotides length. For this purpose full length [^{32}P]- $\Omega\beta$ RNA was digested with low amounts of nuclease S1 as described in fig.3 and in section 2.2.1.1. (the amounts used were found suitable in pre-

Figure 2: Outline of the Construction of Hybrid Plasmids Containing Partial Copies of the $\Omega\beta$ Genome

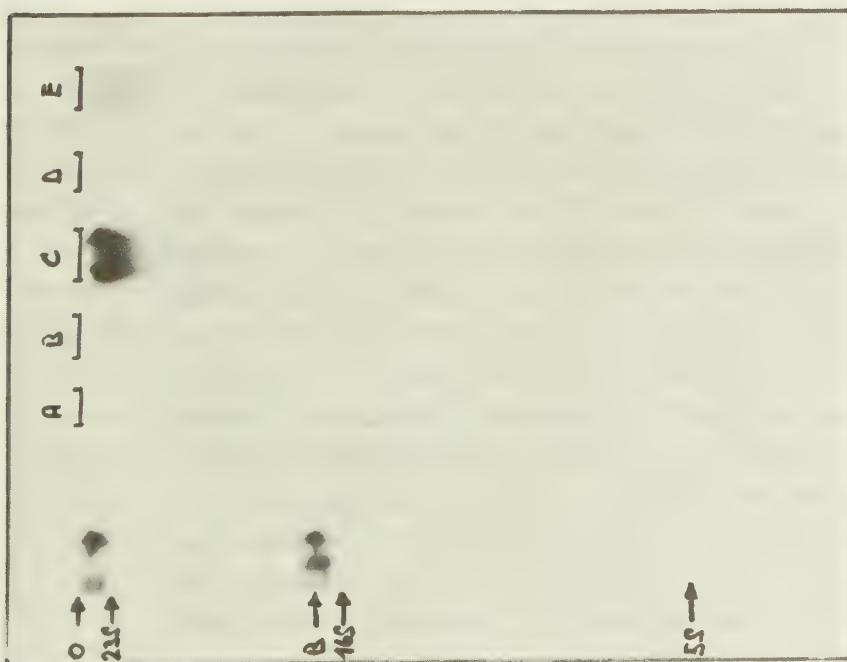


Full size $\Omega\beta$ plus strand RNA is randomly cleaved with nuclease S1 (a). The fragments are elongated with AMP residues by terminal riboadenylate transferase (b) and used as templates for reverse transcriptase, with oligo(dT) as primer (c). The RNA is digested away from the RNA-DNA hybrids with alkali, leaving minus strand cDNAs (d) which are converted to DNA duplexes by an unprimed reverse transcriptase reaction (e); the hairpin loops are now cleaved with nuclease S1 (f). Plasmid PCRI is cleaved with EcoRI and the ends are trimmed back with 5'-exonuclease (h). The double stranded $\Omega\beta$ DNA and the PCRI DNA are elongated with dAMP and dTMP residues, respectively (g,g') with terminal nucleotidyl transferase. The two components are hybridized and transfected into *E.coli* (i).

Figure 3: Polyacrylamide Gel Electrophoresis of Fragmented $\varrho\beta$ RNA

Two portions of 30S $\varrho\beta$ RNA (210 μg each, containing 5×10^5 cpm of ^{32}P) were digested with 0.003 U (lane E) or 0.02 U (lane D) of nuclease S1 for 10 min at 37°C and deproteinized as described in section 2.2.1.1. Analytical aliquots containing 10 μg of RNA were electrophoresed through a 3% polyacrylamide gel (38x18x0.2 cm) in 80 mM Tris-borate (pH 8.3), 7 M urea, 1 mM EDTA at 7.5 V/cm (see section 2.2.4.5.), and the gel was autoradiographed.

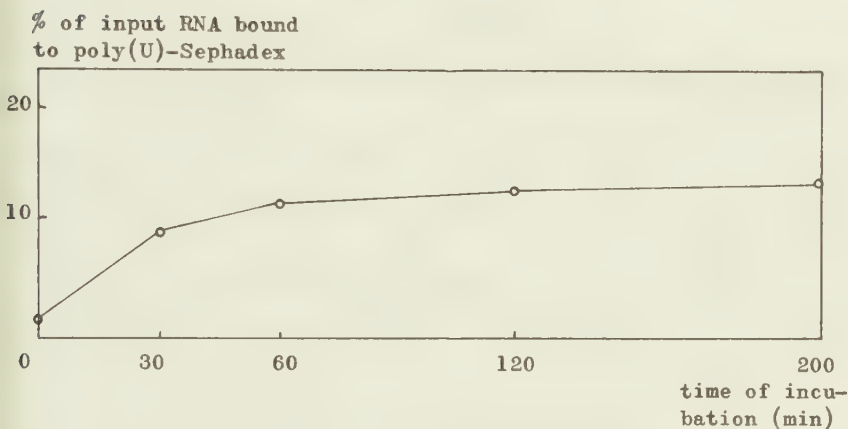
Lane A: unlabelled rRNA of E.coli (5S, 16S, 23S), visualized separately after ethidium bromide staining (section 2.2.4.5.). Lane B: poly(A) 30S $\varrho\beta$ RNA (2 μg , 1000 cpm of ^{32}P); lane C: 30S $\varrho\beta$ RNA (10 μg , 5000 cpm of ^{32}P); 0: origin; B: bromophenol blue dye marker.



liminary experiments; results not shown). For this cleavage nuclease S1 rather than degradation by alkali or other endonucleases was chosen in order to obtain RNA fragments with the free 3'-hydroxyl ends required for elongation with AMP residues by terminal riboadenylate transferase (under mildly alkaline conditions mainly 2',3'-cyclic phosphates are formed, which cannot even be cleaved nor dephosphorylated by phosphatase). An aliquot of the partially degraded RNA was then analyzed by gel electrophoresis. The autoradiograph shown in fig.3 reveals that under these conditions fragments of a broad variety of lengths were formed, with sizes ranging from about 100 to 1000 nucleotides. Only the preparation analyzed in lane E of fig.3 was used for the subsequent steps.

The fragment preparation was now to be elongated at the 3'-ends with AMP residues by terminal riboadenylate transferase (Gilvarg et al., 1975). To find suitable experimental conditions for the elongation reaction the preliminary experiment shown in fig.4 was carried out. It can be seen that after 120 min virtually no further addition of AMP residues seems to take place and that only a low percentage of the fragments contained a sufficiently long poly(A) tail. In the preparative experiment, shown in table 1, incubation was therefore carried out using a double amount of terminal transferase (12 ug of enzyme per ug of RNA) and for a longer time (180 min). Binding of the RNA fragments to poly(U) Sephadex was determined as indicated in the legend of table 1. Fragments which were not bound at all or which were washed off in high salt buffer were taken to represent unspecific or weak binding due to short poly(A) tails. Fragments eluting with formamide buffer represented the fraction of RNA having a large enough number of 3'-terminal adenylate resi-

Figure 4: Time Course of Terminal Riboadenylate Transferase
Reaction



Q β RNA fragments (10 μ g, containing 22'000 cpm of 32 P) were incubated with 60 μ g of terminal riboadenylate transferase and ATP in a total volume of 25 μ l at 37°C as described in section 2.2.1.2. At the time points indicated 2 μ l-aliquots were withdrawn from the reaction, cooled to 0°C and diluted into 100 μ l of 1 mM EDTA. After heat denaturation (100°C for 90 sec) and quenching in ice-water the samples were chromatographed on poly(U)-Sephadex G-10 (0.2 ml bed volume) as described in section 2.2.1.2. The radioactivity of the fractions was determined by measuring Cherenkov radiation. The percentages shown indicate the fraction of RNA eluted from the columns with 90% formamide buffer.

Table 1: Binding of ^{32}P - $\text{Q}\beta$ Fragments Elongated with 3'-Poly(A)
Tails to Poly(U)-Sephadex G-10

	pA- $\text{Q}\beta$ RNA fragments		control fragments	
	cpm	%	cpm	%
flowthrough and elution with 0.5 M NaCl, 10 mM Tris-HCl (pH 7.5) 0.5% SDS	45'320	71.5	5360	99.3
elution with 90% formamide, 10 mM Tris-HCl (pH 7.5), 0.5% SDS	16'480	26.0	25	0.5
residue not eluted from column	1590	2.5	10	0.2

$\text{Q}\beta$ RNA (200 μg of SI-fragments, containing 6×10^4 Cherenkov cpm) was incubated with ATP and terminal riboadenylate transferase (2.4 μg) for 180 min at 37°C in a volume of 500 μl as described in section 2.2.1.2. In the control reaction one tenth the amount of $\text{Q}\beta$ RNA fragments was incubated under the same conditions, omitting the enzyme. The reaction was stopped by adding EDTA (20 mM final concentration) and cooling to 0°C . After deproteinization, precipitation and centrifugation (see section 2.2.1.1.) the RNA pellets were dissolved in 100 μl of 1 mM EDTA, denatured for 45 sec at 100°C , quenched in ice-water and the mixture was adjusted to a final concentration of 0.5 M NaCl, 10 mM Tris-HCl (pH 7.5), 0.5% SDS. Samples were adsorbed to columns of poly(U)-Sephadex G-10 (0.2 ml bed volume), and the columns were eluted with 1 ml each of the buffers indicated (see section 2.2.1.2.). The radioactivity of the eluted fractions was determined by measuring Cherenkov radiation. Percentages indicate percent of total Cherenkov cpm.

dues added to allow successful hybridization of oligo(dT) primer molecules for reverse transcription. RNA from this fraction was therefore precipitated and used in the subsequent steps.

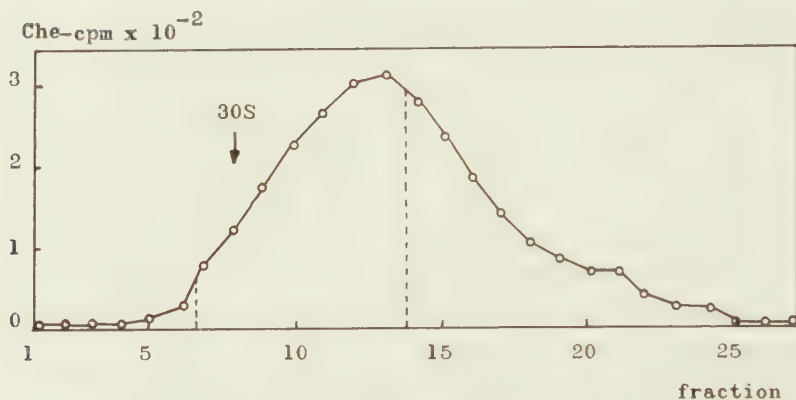
3.1.2. Synthesis of Double Stranded DNA Containing Sequences of $\alpha\beta$ RNA

The preparation of poly(A) elongated RNA fragments was anticipated to contain still an appreciable amount of small (<10S) RNA fragments. As this small RNA, present in great molar excess over the larger RNA fragments, would have resulted in the formation of large numbers of hybrid plasmids containing only small DNA transcripts the small RNAs were removed by centrifugation through a 5-23% sucrose gradient, and only the large RNA fragments were used in subsequent steps (see fig.5).

For the synthesis of $\alpha\beta$ cDNA poly(A) elongated $\alpha\beta$ RNA fragments (5 μg) were incubated with AMV reverse transcriptase in the presence of oligo(dT)₁₂₋₁₈ primer and α -[³²P]-dGTP as described in section 2.2.1.3. 1.2 μg of cDNA was recovered from the reaction with the $\alpha\beta$ RNA fragments, corresponding to 24% transcription of the input RNA into minus strand cDNA.

Synthesis of the plus strand DNA was carried out in an unprimed reaction with reverse transcriptase: it had been found (Marniatis et al., 1976) that upon removal of the RNA from the RNA-DNA hybrid (e.g. by alkali) the 3'-end of the single stranded cDNA would fold back on itself creating a hairpin that would serve as primer-template for the synthesis of the plus strand DNA (see scheme in fig.2). The synthesis of the second DNA strands was carried out with 1.2 μg of ³²P-labelled cDNA, in principle under the same conditions used for the synthesis of the first DNA strands. The radioactive substrate used in this

Figure 5: Centrifugation of Poly(A) Elongated $Q\beta$ RNA Through 5-23% Sucrose Gradients



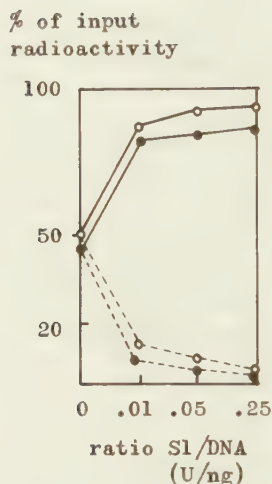
Poly(A) elongated ^{32}P - $Q\beta$ RNA fragments (52 μg , 3000 Cherenkov cpm) were centrifuged through a 5-23% sucrose gradient in a Spinco SW-60 rotor at 55'000 rpm and 15°C for 150 min, as described in section 2.2.1.2. The radioactivity of the fractions (0.165 ml/fraction) was determined by measuring Cherenkov radiation. Fractions between the two dashed vertical lines were pooled and the RNA was precipitated as described in section 2.2.1.1. 44% (23 μg , containing 1320 Cherenkov cpm) of the input RNA was recovered in the precipitate.

The arrow shown corresponds to 30S $Q\beta$ RNA centrifuged as a size marker in a parallel tube of the gradient.

reaction was [^3H]-dATP (specific activity 1.6×10^5 cpm/nmol; for experimental details see section 2.2.1.4.). In this way the first (minus) DNA strand carrying a ^{32}P -label could be differentiated from the second (plus) strand, labelled by [^3H]-dAMP. In the plus strand DNA reaction 375 pmol of [^3H]-dAMP were incorporated, corresponding to 45% transcription of the first (minus) DNA strands (0.5 μg newly synthesized plus strand DNA).

To obtain normal double stranded 2β DNA the terminal hairpincovalently linking the two strands had to be cleaved by nuclease S1. The amount of nuclease to cleave the terminal loop, yet to keep the double stranded region unnicked, had to be determined in a preliminary experiment. Fixed amounts of the double stranded DNA preparation were incubated with varying amounts of nuclease S1 as described in section 2.2.1.5. To distinguish between cleaved and uncleaved molecules the preparations were heat-denatured, chilled quickly and diluted: only molecules in which the two complementary strands are still connected covalently by the hairpin loop will renature, whereas the strands of cleaved molecules will stay separate. The samples were then chromatographed on hydroxyapatite columns, which served to differentiate between single- and double-stranded DNAs, since double stranded molecules are more tightly bound to the column than single strands. The results of the analysis on hydroxyapatite are shown in fig.6. It is interesting to note that even without any nuclease treatment almost 50% of the DNA preparation was eluted in the single stranded fraction after denaturation, indicating a possible contamination of the DNA preparation or the reverse transcriptase with a nuclease. Fig.6 shows that as little as 0.01 U of S1/ng of DNA was sufficient to convert the DNA completely to denaturable double strands.

Figure 6: Assay of Single- and Double-Stranded DNA after Treatment with Nuclease S1 by Chromatography on Hydroxyapatite



Q β DNA (10 ng, containing 650 cpm of ^{32}P and 350 cpm of ^3H) was incubated with varying amounts of nuclease S1 (0, 0.1, 0.5, 2.5U) in a total volume of 10 μl as described in section 2.2.1.5. The mixture was adjusted to a final concentration of 2 mM Tris-HCl (pH 7.5), 4 mM EDTA, 0.4% SDS in a total volume of 20 μl . The DNA was heat-denatured (2 min at 100°C) and quickly chilled in ice-water. 1.0 ml of 0.14 M phosphate buffer (pH 6.8) was added to the mix-

ture and the preparation was chromatographed on a hydroxyapatite column (0.2 ml bed volume) as described in section 2.2.1.5. 0.14 M phosphate buffer eluted single stranded DNA, whereas 0.4 M phosphate buffer eluted the double stranded fraction. The DNA in the eluted fractions was precipitated with cold 6% TCA in the presence of 20 μg of yeast carrier RNA, filtered through millipore discs and the ^3H - and ^{32}P -radioactivity was determined by scintillation counting.

Minus strand DNA (^{32}P) in single strands (●—●) and double strands (●---●); plus strand DNA (^3H) in single strands (○—○) and double strands (○---○).

To evaluate the intactness of the double stranded $\mathcal{Q}\beta$ DNA after treatment with nuclease S1 an aliquot of the DNA was sedimented through a sucrose gradient under denaturing conditions before and after treatment with nuclease S1, as shown in fig.7. The experiment shows that under the conditions used nuclease S1 cleaves only the hairpin loop of the double stranded $\mathcal{Q}\beta$ DNA, leaving the double stranded portion proper mostly intact. It is also seen that the bulk of the $\mathcal{Q}\beta$ DNA sediments faster than the globin cDNA used as a size marker (approximately 600 nucleotides).

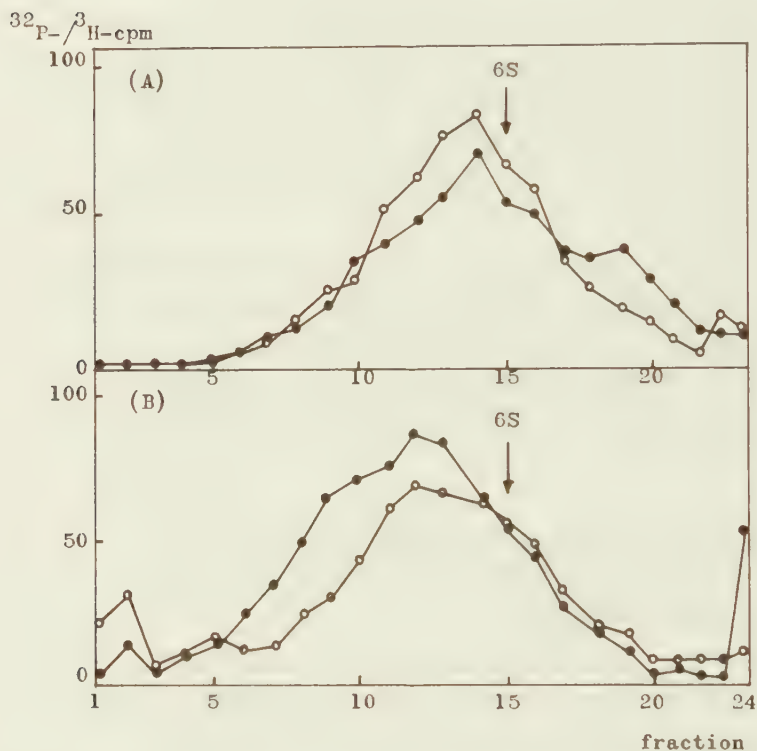
In the preparative reaction 1.3 μg of double stranded hairpin loop DNA was cleaved with nuclease S1 as described in section 2.2.1.5., yielding a product that could be denatured to 90% and which, after denaturation, did not differ appreciably in its size from a sample not treated with nuclease S1 (results not shown).

3.1.3. Elongation of Double Stranded $\mathcal{Q}\beta$ DNA with Poly(dA), Annealing with Poly(dT)-Elongated Plasmid PCRI and Transfection of E.coli

A portion of the nuclease treated double stranded $\mathcal{Q}\beta$ DNA was elongated at the 3'-termini with poly(dA) by incubation with terminal deoxyribonucleotide transferase in the presence of dATP, as shown in the legend of table 2. To analyze the completeness of this reaction half of the preparation was chromatographed on poly(U)-Sephadex G-10 as described in section 2.2.1.2.

Table 2 shows that most of the DNA strands (about 65%) had obviously been elongated with dA-residues. For this reason prefractionation of the DNA on poly(U)-Sephadex was thought to be unnecessary, and the $\mathcal{Q}\beta$ DNA preparation was used directly for the further reaction.

Figure 7: Sucrose Gradient Centrifugation of the $\Omega\beta$ Double Stranded DNA Before and After Treatment with Nuclease S1



Double stranded $\Omega\beta$ DNA (25 ng, containing 1000 cpm of ^3H and 900 cpm of ^{32}P) was treated with nuclease S1 (0.25 U; 0.01 U per ng of DNA) (panel A) or left untreated (panel B) as described in section 2.2.1.5. After pronase/SDS/phenol treatment and chromatography on Sephadex G-100 salmon sperm DNA (20 μg) was added as carrier and the DNA was precipitated as described in section 2.2.1.1.

The precipitates were dissolved in 20 μl of 1 mM EDTA, heat-denatured at 100°C for 3 min and the preparations were quickly

chilled in ice-water. The mixtures were then adjusted to a final concentration of 0.2 M NaOH, 0.1% SDS (final volume 100 μ l) and layered on a 5-23% sucrose gradient containing 0.2 M NaOH, 0.8 M NaCl, 0.01% SDS (4 ml gradient volume). The samples were centrifuged in a Spinco SW-60 rotor at 15°C and 36'000 rpm for 14 h. The gradients were fractionated from the bottom, and fractions (0.13 ml) were collected on paper filter discs (15 mm diameter). After drying at 30°C for 2 h the radioactivity on the filter discs was determined by liquid scintillation counting.

The arrows in panels A and B indicate the position of 6S mouse globin cDNA (gift of Dr.H.Hofstetter), run under the same conditions as a size marker in two parallel gradient tubes.

Minus strand DNA (32 P) is indicated by (●—●), plus strand DNA (3 H) by (○—○).

Table 2: Chromatography of $\phi\beta$ DNA on Poly(U)-Sephadex G-10 after Reaction with Terminal Nucleotidyl Transferase and Denaturation

	$\phi\beta$ plus strand		$\phi\beta$ minus strand	
	^3H -cpm	%	^{32}P -cpm	%
flowthrough and elution with 0.5 M NaCl, 10 mM Tris-HCl (pH 7.5) 0.5% SDS	220	28.2	100	17.3
elution with 90% formamide, 10 mM Tris-HCl (pH 7.5), 0.5% SDS	490	62.8	400	68.9
residue not eluted from column	70	9.0	80	13.8

Double stranded $\phi\beta$ DNA (80 ng, containing 3000 cpm of ^3H and 2700 cpm of ^{32}P) was incubated with terminal deoxyribonucleotide transferase (0.5 U) in the presence of dATP (0.1 mM), as described in section 2.2.1.6. After deproteinization, addition of 0.2 μg of ϕX174 carrier DNA and chromatography through Sephadex G-100 (see section 2.2.1.6. for experimental details) 44% of the input DNA was recovered (35 ng; 1320 cpm of ^3H , 1200 cpm of ^{32}P). Half of this preparation (17 ng, containing about 700 cpm of ^3H and 600 cpm of ^{32}P in a total volume of 100 μl of 10 mM Tris-HCl (pH 7.5), 0.1 mM EDTA) was diluted to 250 μl of 0.2% SDS, 10 mM Tris-HCl (pH 7.5) and 0.8 mg/ml yeast RNA. After heat-denaturation the sample was processed and chromatographed on poly(U)-Sephadex G-10 as described in the legend of table 1. After addition of yeast RNA (100 μg /fraction) the DNA in each fraction was precipitated onto Millipore filters with cold TCA (6%), and the radioactivity in the fractions was determined by scintillation counting. All values shown are corrected for background cpm (^3H : 18 cpm, ^{32}P : 35 cpm); percentages are fraction of total cpm.

In the next step the poly(dA) elongated $\text{Q}\beta$ DNA ($\text{Q}\beta$ DNA-pA) was annealed in varying amounts to samples of plasmid PCRI DNA which had been cleaved at its unique EcoRI restriction site and had been elongated with dTMP-residues at the 3'-termini (PCRI-pT). This preparation was then used to transfect *E.coli*, as shown in table 3.

E.coli clones transfected with PCRI plasmid (with or without insert DNA) were selected by their resistance to kanamycin conferred by the km^r -gene in plasmid PCRI. The bacterial strain used in this experiment is dependent on streptomycin to prevent these bacteria (transformed or not transformed) from growing outside the laboratory.

Table 3 shows that for the $\text{Q}\beta$ -PCRI hybrid plasmid preparations (plates 2 to 4) the specific infectivity is approximately inversely proportional to the amount of $\text{Q}\beta$ DNA-pA used in the experiments. This may indicate an inhibitory compound in the $\text{Q}\beta$ DNA-pA preparation, interfering with the annealing or the transfection steps. Specific infectivities of $\text{Q}\beta$ DNA-PCRI hybrids are of the same order of magnitude as the infectivity of the globin DNA-PCRI hybrid (plate 5). The parental plasmid PCRI shows at least a 200-fold higher infectivity (plate 6) than any derived DNA-PCRI hybrid preparations, indicating a rather low efficiency of the annealing and in vivo ligation steps.

3.1.4. Identification of Bacteria Harbouring Hybrid Plasmids with $\text{Q}\beta$ Specific Sequences and Preliminary Classification of Selected Plasmids

To identify bacteria harbouring hybrid plasmids which contained $\text{Q}\beta$ specific sequences 240 bacterial colonies from the transfection experiment described in section 3.1.3. were chosen at random, inoculated into 2 ml of normal medium containing kanamycin and streptomycin and grown to 10^8 cells/ml (see 2.2.4.1.).

Table 3: Transfection of E.coli N543 with $\alpha\beta$ DNA-PCR1 Hybrid
Plasmid Preparation

plate no.	PCR1-pT (ng)	PCR1-fI (ng)	$\alpha\beta$ DNA-pA (ng)	globin DNA-pA (ng)	colonies/plate	colonies/ng of insert DNA ^(*)
1	10	-	-	-	63	(*)
2	10	-	4.0	-	207	36
3	10	-	1.0	-	189	189
4	10	-	0.5	-	300	474
5	10	-	-	0.5	223	320
6	-	0.5	-	-	$\sim 5 \times 10^3$	-

For the annealing step prior to transfection a fixed amount of PCR1 DNA-pT (10 ng, 1.1 fmol; preparation B148C, gift from Dr. J.van den Berg) and $\alpha\beta$ DNA-pA (0.5 ng, 1 ng, 4 ng; 0.75 fmol, 1.5 fmol, 6 fmol, respectively; number of moles calculated assuming an average size of 1000 base pairs for $\alpha\beta$ DNA, see fig.7) were mixed in 25 μ l of TNE (plates 2 to 4). The following control transfections were carried out: PCR1 DNA-pT alone (plate 1), equimolar amounts of PCR1 DNA-pT and rabbit globin cDNA-pA (preparation DS215; 0.5 ng, 1.25 fmol, assuming a size of 600 base pairs for globin DNA; plate 5), uncleaved form I PCR1 DNA (PCR1-fI, 0.5 ng, plate 6). The preparations were annealed, transfected into E.coli N543 and plated on agar containing kanamycin and streptomycin as described in section 2.2.2.2. After 18 h at 37°C (no further increase in number of colonies) the kanamycin-resistant colonies (0.5-2 mm diameter) were counted.

(*) For calculating the specific infectivity the number of colonies found in plate 1 (PCR1-pT background) was subtracted from every other colony count; the colonies found in plate 1 were probably due to uncleaved form I PCR1 in the PCR1-pT preparation, no specific infectivity is defined here.

Small aliquots of each culture were transferred to Millipore filter discs lying on agar, the colonies were grown up and the filters were processed for the Grunstein-Hogness filter hybridization assay as described in section 2.2.3.

Several types of ^{32}P -labelled probes were used in the filter hybridization assays: a probe for the complete genome of $\text{Q}\beta$ phage and probes for selected, limited regions of $\text{Q}\beta$ RNA.

In the experiment with the full-length $\text{Q}\beta$ probe 60 out of the 240 clones chosen did not show hybridization (this percentage was to be expected from the background found in the transfection experiment; see table 3. plate 1: self-ligated PCRI-pT or residual, uncleaved form I PCRI). 61 clones showed a weak response to this probe, and 119 clones showed moderate to strong hybridization with the full-length ^{32}P - $\text{Q}\beta$ probe (see fig.8, panel (a): hybridization with ^{32}P - $\text{Q}\beta$ plus strand RNA to clones p $\text{Q}\beta$ 121-240).

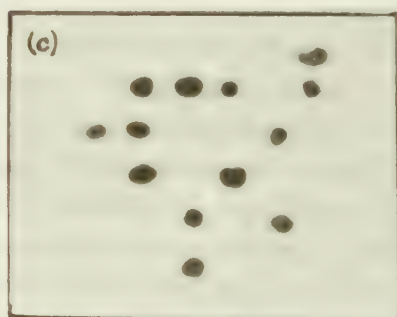
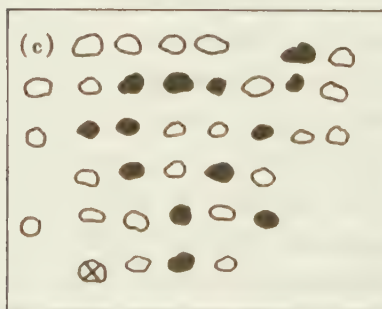
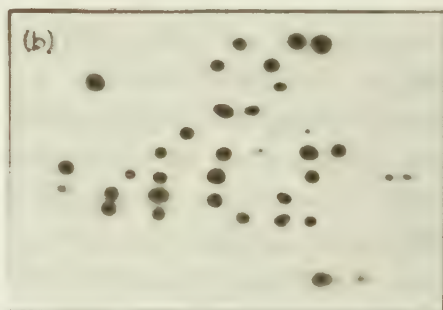
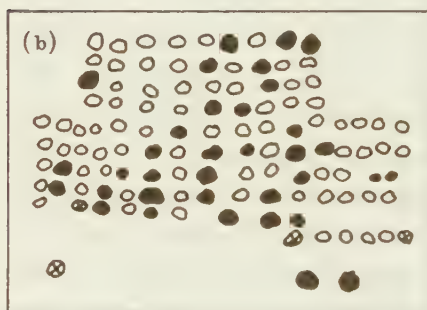
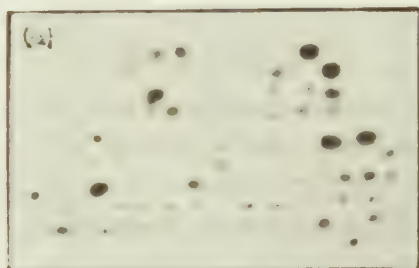
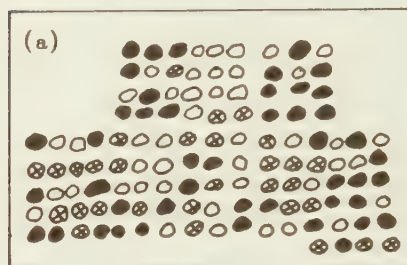
Hybridization with smaller probes of known location in the $\text{Q}\beta$ genome (Billeter, 1978) allowed to determine approximately which region of the $\text{Q}\beta$ sequence was present in the hybrid plasmid of each bacterial clone. Probes used in this experiment were RNase T_1 -resistant oligonucleotides (T1^+ , T2^+ , T5^+ and T6^+ from $\text{Q}\beta$ plus strand RNA; T1^- , T3^- , T7^- and T24^- from $\text{Q}\beta$ minus strand RNA), fragments M-2, M-11 and S-3 from the replicase binding sites (gift of Dr.F.Meyer) and probes specific for the 3'- and 5'-terminal regions of the $\text{Q}\beta$ genome. In fig.8 the results for colony hybridization with two of these site-specific probes are shown: 3'-end probe (panel (b)) and oligonucleotide T1^- (panel (c)).

The preliminary classification of $\text{Q}\beta$ partial genome copy plasmids on the basis of hybridization with site-specific probes allowed to single-out certain hybrid plasmids for the determination of characteristic restriction sites within the $\text{Q}\beta$

Figure 8: Identification and Characterization of $\text{Q}\beta$ Plasmids
by Colony Filter Hybridization

(tracing of colonies on filter)

(autoradiography)



(hybridization response: strong (●), weak (⊗), negative (○))

(legend fig.8)

Bacterial colonies were grown on nitrocellulose filters and hybridizations with the radioactive probes were carried out as described in section 2.2.3.

Panel (a) shows hybridization of ^{32}P - $\text{Q}\beta$ RNA (full-length plus strands, specific activity 1.5×10^6 cpm/nmol) to clones p $\text{Q}\beta$ 121 to 240. The different intensities of blackening are probably due to the variable lengths of the $\text{Q}\beta$ specific sequences inserted in the hybrid plasmid.

Panel (b) shows hybridization of a ^{32}P -labelled minus strand RNA probe complementary to the 3'-end of $\text{Q}\beta$ plus strand RNA to clones p $\text{Q}\beta$ 121-240. The probe was prepared by first incubating 20 μg of $\text{Q}\beta$ full-length plus strand RNA (preparation V-84) with 50 U of $\text{Q}\beta$ replicase (preparation TW352) and 60 μg of HFI (preparation TW327) in a total volume of 250 μl containing rATP (0.8 mM), rGTP (1.0 mM), Tris-HCl (pH 7.5; 80 mM), MgCl_2 (12 mM) and EDTA (1 mM) for 5 sec at 37°C . The G- and A-containing tetradecanucleotide (Bandle, 1973) was then elongated further by adding rCTP (0.5 mM) and α - ^{32}P -rUTP (0.05 mM, specific activity 1.2×10^7 cpm/nmol), and the mixture was incubated for further 15 sec at 37°C . The reaction was stopped by the addition of EDTA (30 mM final concentration), and after pronase/SDS/phenol treatment the RNA was precipitated with ethanol and TNE (see section 2.2.1.1.). Total yield was 2.3×10^6 cpm of ^{32}P , corresponding to 3.3 μg of newly synthesized minus strand RNA fragments, with an average chain length of approximately 250 nucleotides. The minus strand fragments were separated from the full-length $\text{Q}\beta$ template RNA by heat-denaturation and centrifugation in a 5-23% sucrose gradient (Spinco SW-60 rotor, for 16 h at 4°C and 42'000 rpm), and fractions corresponding to 5-8S sedimentation coefficient were collected (results not shown). RNase A resistance measurements showed a plus strand content of 1.4% in the minus strand fragment

(legend fig.8)

preparation (results not shown). This probe was used for hybridization to the filters as described in section 2.2.3.

(specific activity 5×10^6 cpm/nmol minus strand fragment).

Panel (c) shows the hybridization with $\Delta\beta$ minus strand oligonucleotide T1⁻. This oligonucleotide had been prepared by digestion of ^{32}P - $\Delta\beta$ minus strand RNA with ribonuclease T₁, separation of the oligonucleotides by two-dimensional gel electrophoresis and preparative isolation of the oligonucleotide(s) from the gel (De Wachter & Fiers, 1972). The specific activity of this probe was 5×10^6 cpm/nmol oligonucleotide.

sequence. Fig.9 summarizes the outcome of such hybridization and restriction experiments for various selected plasmids: it can be seen that in the ensemble of plasmids PQ β 61, 71, 74, 76, 100, 141 and 160 the whole sequence of the Q β genome is represented (except for the actual 5'- and 3'-ends of the genome; due to the large size of the hybridization probes specific for these regions (about 200 nucleotides) and for the lack of suitable restriction sites at the actual ends of the genome it was not possible to delimit exactly the end points of the Q β specific sequences in plasmids PQ β 160 and 74, respectively).

To prove that the plasmids indeed contained Q β specific sequences flanked by two poly(dA)-poly(dT) regions and to estimate the length of the insert sequences several hybrid plasmids were digested with low amounts of nuclease S1 under partially denaturing conditions (Hofstetter et al., 1976). The fragments obtained were separated by electrophoresis in agarose gels (see section 2.2.4.5.) and the fragment containing Q β specific sequences was identified by hybridization with ³²P-labelled Q β RNA, using the Southern blot technique (Southern, 1975).

Fig.10 shows this type of experiment for plasmid PQ β 71. The excised Q β specific DNA fragment has a size of approximately 1800 base pairs, corresponding to approximately 40% of the total length of the Q β genome (for the location of this sequence on the Q β genome see fig.9). This identification of insert sequences was carried out for a series of plasmids: an insert length of 1000 nucleotides was estimated for plasmid PQ β 61, 1200 nucleotides for PQ β 74, 1100 nucleotides for PQ β 76 and 1300 nucleotides for PQ β 160 (results not shown).

Since it was clear that most probably the whole nucleotide sequence of the Q β genome was represented in the plasmid collec-

(legend fig.9)

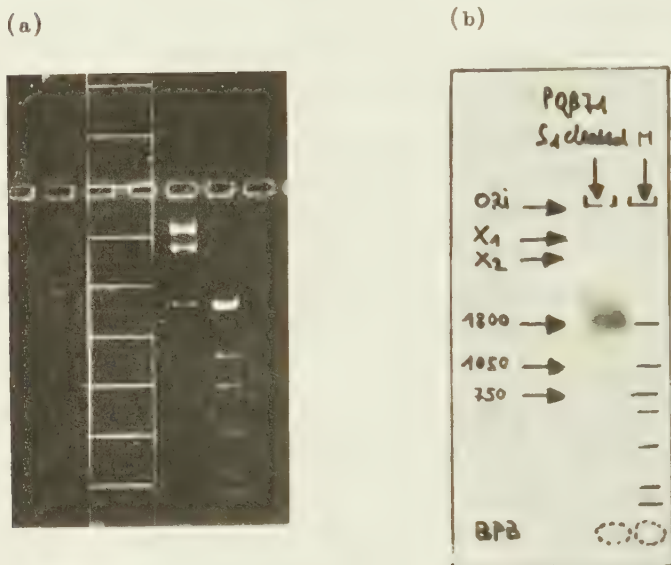
Restriction sites are indicated by (↓) on the map, (↔) denotes the actual experimental evidence of the restriction site in the nucleotide sequence; hybridization probes are shown as (←) on the map, (■) means actual experimental evidence for the hybridization with the probe.

Hybridizations were carried out according to the Grunstein-Hogness procedure, as described in section 2.2.3. and as documented in fig.8. ^{32}P -labelled probes were the T_1 -oligonucleotides T1^+ , T2^+ , T5^+ , T6^+ from the plus strand RNA, and T1^- , T3^- , T24^- from $\text{Q}\beta$ minus strand RNA. A second class of somewhat larger probes were ^{32}P -labelled replicase binding fragments M-2 and S-3 (gift of Dr.F.Meyer), or ^{32}P -labelled probes corresponding to the 5'- and 3'-end of $\text{Q}\beta$ plus strand RNA (for details see fig.8, legend).

For restriction site analysis only restriction endonucleases specific for hexanucleotide recognition sequences were used. Plasmids for the restriction enzyme assays were isolated from the clones indicated using the small-scale isolation protocol described in section 2.2.4.2. The various $\text{Q}\beta$ hybrid plasmids and the parental plasmid PCRI (typically 0.3-0.5 μg) were digested in parallel reactions with the restriction enzyme of interest (see section 2.2.4.4.), and the cleavage products were analyzed by electrophoresis in the analytical agarose-gels described in section 2.2.4.5. (gels not shown).

The map shown here was constructed by arranging the $\text{Q}\beta$ specific sequences of the various plasmids according to the known location of the site-specific hybridization probes (Billeter, 1978). Results from restriction cleavage analysis served to support the ordering of the various sequences and allowed the construction of a first approximate restriction map of the $\text{Q}\beta$ genome. The distances shown are the actual distances as evidenced by nucleotide sequence determination.

Figure 10: Nuclease S1-Excision of the DNA Insert in Hybrid Plasmid PQ β 71: Identification as a Q β Specific Sequence and Estimation of its Size



Plasmid PQ β 71 DNA (1 μ g) was digested with nuclease S1 (3 U) at 50°C for 30 min in a total volume of 20 μ l containing 50% formamide buffer as described in section 2.2.4.3. (the optimal formamide concentration was determined in a preliminary experiment; results not shown). The preparation was then electrophoresed in a 2% agarose gel (9.5x8.0x0.2 cm), the gel was stained with ethidium bromide and photographed under UV-light as described in section 2.2.4.5. (panel (a), left lane). Double stranded DNA fragments of known sizes were run in parallel (panel (a), right lane). The size markers are from plasmid P β G DNA digested with restriction endonuclease EcoRI and a combination of endonucleases EcoRI and HhaI. All sizes indicated are approximate and are given in numbers of base pairs. The

(legend fig.10, continued)

numbers have been derived by coelectrophoresis with ϕ X174 DNA (the sequence of which is known; Sanger et al., 1977) cleaved with restriction endonuclease HaeIII.

The DNA in the agarose gel was transferred to a sheet of nitrocellulose and hybridization with ^{32}P -labelled $\Omega\beta$ plus strand RNA (6×10^6 cpm of ^{32}P , specific activity 8×10^6 cpm/ μg) was carried out according to the method of Southern (Southern, 1975). Panel (b) shows the autoradiogram of the Southern-blot of the gel shown in panel (a) (the location of the marker fragments are indicated). In the left lane of panel (b), S1-treated plasmid P $\Omega\beta$ 71 preparation shows strong hybridization with the $\Omega\beta$ probe for a fragment moving to a position corresponding to the 1800 base pair size marker fragment: this 1800 base pair fragment hybridizing with $\Omega\beta$ RNA is the actual $\Omega\beta$ specific insert of plasmid P $\Omega\beta$ 71. X_1 is the linear form (form III) of the entire hybrid plasmid P $\Omega\beta$ 71 (only one of the poly(dA)-poly(dT) regions cleaved, therefore hybridization with the $\Omega\beta$ probe); X_2 contains only the parental plasmid PCRI (both poly(dA)-poly(dT) regions cleaved, therefore no hybridization with the $\Omega\beta$ probe due to its lack of $\Omega\beta$ sequences).

tion, nucleotide sequence determination was then started. First, the vicinity of the unique EcoRI site and the 5'-terminal XhoI site was investigated with the aim to establish the whole nucleotide sequence of the Al-cistron, a large part of which is known from RNA sequencing studies (see section 4.1.).

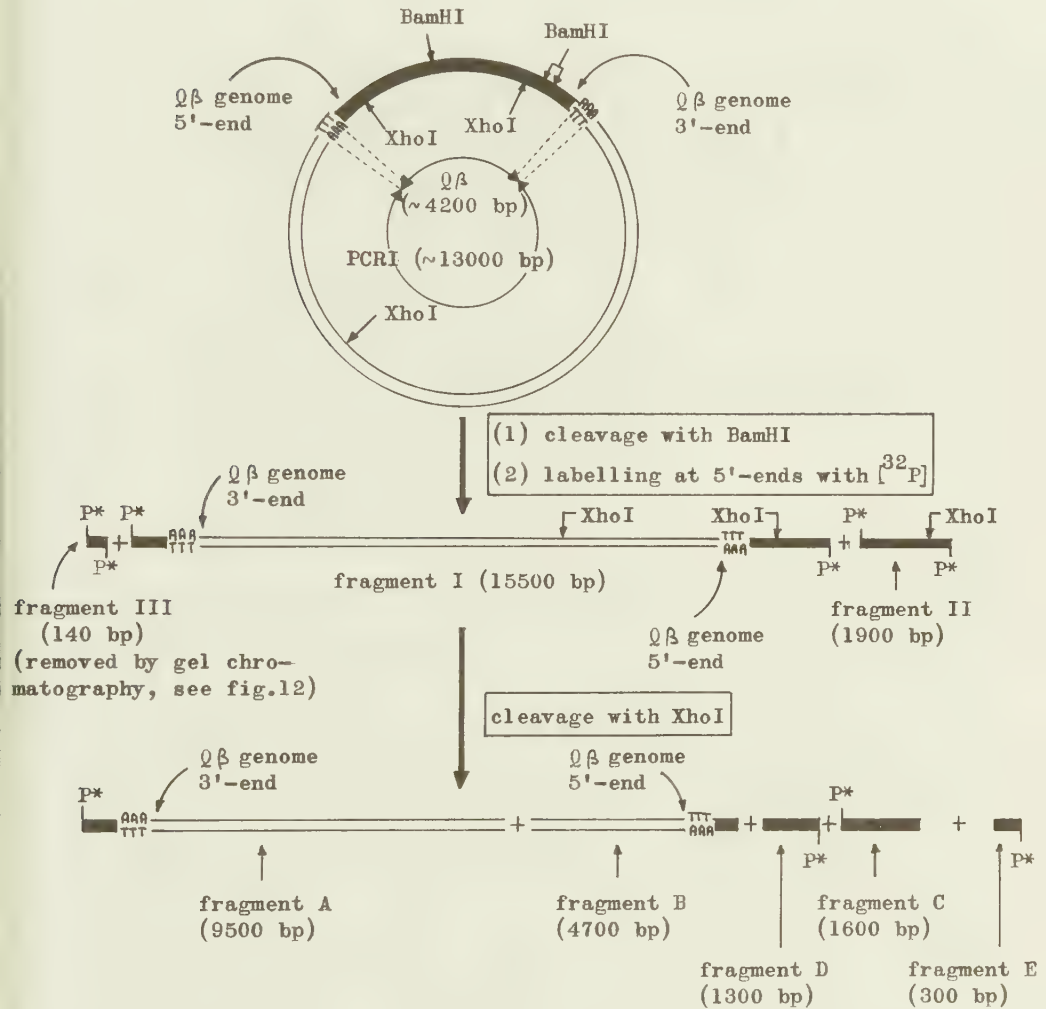
3.2. Restriction Map of the $\Omega\beta$ Genome

For this part of the work the plasmid Zp $\Omega\beta$ 32 was used, which was constructed some time after the isolation of the plasmids described above and which contained a complete copy of the $\Omega\beta$ genome (Taniguchi et al., 1978).

For restriction mapping the methodology described by Smith and Birnstiel (Smith & Birnstiel, 1976) was adopted. The strategy followed in our case was as outlined in fig.11. The $\Omega\beta$ sequence in plasmid Zp $\Omega\beta$ 32 was first cleaved with restriction endonuclease BamHI. The fragments were labelled at their 5'-termini with [32 P]-phosphate, yielding molecules labelled at both 5'-ends of the double stranded DNA. Cleavage of the fragments with a second restriction enzyme which cuts each fragment only at one position (here endonuclease XhoI) allowed the isolation of fragments carrying [32 P]-phosphate groups only at one of the 5'-ends (see fig.11).

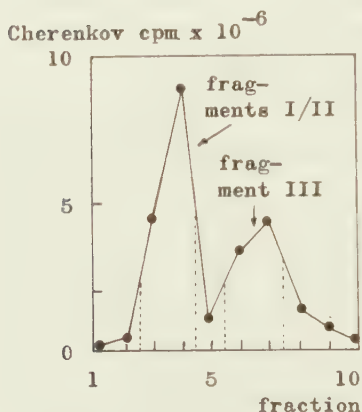
The actual isolation of fragments A, C, D, E (see fig.11) used for restriction mapping involved in a first step the removal of the small fragment III by gel chromatography (see fig.12). In a second step the large fragments I and II were cleaved with restriction endonuclease XhoI and the digestion products were resolved by agarose gel electrophoresis (fig. 13). The isolated fragments were now treated with the restriction enzymes of interest under various conditions allowing very partial to complete digestion. The length of the digestion

Figure 11: Scheme of Generation of Restriction Fragments
Labelled at One End for Restriction Mapping



Qβ specific sequences are shown as (■), PCRI sequences as (□).
Fragments I, II, III are documented in fig.12; fragments A, B, C, D, E are shown in fig.13. Only fragments A, C, D, E carry a [³²P]-phosphate group (P*) at one 5'-end and are used for restriction mapping; fragment B is not labelled and is not detected in autoradiograms (see fig.13). Fragment sizes are given in base pairs (bp).

Figure 12: Separation of Large Labelled DNA Fragments from Short Fragments by Gel Chromatography

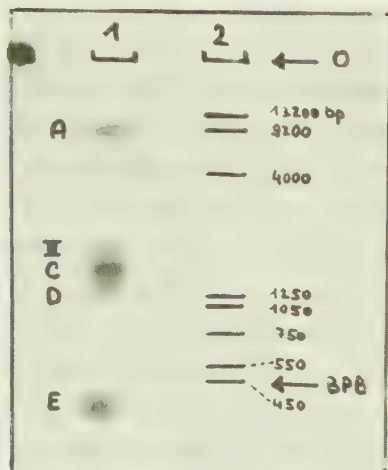


Plasmid ZpQ β 32 DNA (100 μ g) was digested with restriction endonuclease BamHI (100 U) in a total volume of 525 μ l of Y-100 buffer (see section 2.2.4.4.) for 60 min at 37°C. The preparation was passed through Chelex-100, dephosphorylated (see section 2.2.5.), treated with pronase/SDS/phenol and the DNA was precipi-

tated as described in section 2.2.1.1. The dephosphorylated DNA was then labelled by incubation with T4-polynucleotide kinase (10 U) and γ -[³²P]-ATP (80 pmol) in a total volume of 90 μ l at 37°C for 12 min (see section 2.2.5.). After chromatography through Sephadex G-100 (see section 2.2.1.3.) the DNA was precipitated (total yield was 25x10⁶ Cherenkov cpm). The terminally labelled DNA was passed through a Sepharose 4-B column (10 ml bed volume), and fractions of 0.6 ml were collected at a flow rate of 5 ml/h. The radioactivity of the fractions was determined by measuring Cherenkov radiation, and the DNA in fractions 3, 4 (peak containing fragments I and II) and in fractions 6, 7 (peak containing fragment III) was precipitated as described in section 2.2.1.1. Yields were 13.5x10⁶ Cherenkov cpm for fragments I and II, 8x10⁶ Cherenkov cpm for fragment III.

An aliquot of each peak (containing 20'000 Cherenkov cpm) was tested by electrophoresis in an agarose gel (1.2%): the first peak consisted of two fragments (about 15'500 and 1900 base pairs), and the second peak of one fragment (140 base pairs); results not shown, see fig.11 for scheme.

Figure 13: Resolution of ^{32}P -Labelled DNA Fragments by Agarose Gel Electrophoresis



DNA fragments from the cleavage of ZpQ β 32 DNA with restriction endonuclease BamHI (fragments I and II; 60 μg containing 13.5×10^6 Cherenkov cpm, see fig.12) were digested with restriction endonuclease XhoI (10 U) for 120 min at 37°C in 270 μl of Y-150 buffer (see 2.2.4.4.). After pronase/SDS/phenol treatment (see 2.2.1.1.) an aliquot of the preparation (containing 10% of the total) was analyzed by electrophoresis in an 1.2% agarose gel (9.5x8.0x0.2 cm) as described in section 2.2.4.5. Lane 1 in the autoradiogram shows the fragments expected from such a cleavage: fragment A (9500 base pairs), fragment C (1600 base pairs), fragment D (1300 base pairs) and fragment E (300 base pairs). A further fragment migrating to a position which corresponds to approximately 1900 base pairs is seen in addition : by isolation of this fragment and redigestion with endonuclease XhoI this fragment was shown to be identical to

fragment II from the BamHI-cleaved ZpQ Δ 32 DNA (see fig.11), indicating incomplete cleavage with XhoI in this experiment (results from isolation and redigestion experiment not shown). The bulk of the preparation was therefore redigested with XhoI using the conditions given above and the cleavage products were resolved into the component fragments by preparative gel electrophoresis (1.2% agarose, see section 2.2.4.5. for experimental details; results not shown). The fragments in the gel were located by autoradiography, cut out from the gel, eluted and purified as described in section 2.2.4.5. Final yields were (radioactivity expressed in Cherenkov cpm) 1.6×10^6 cpm (fragment A), 3.6×10^6 cpm (fragment C), 1.65×10^6 cpm (fragment D) and 2.2×10^6 cpm (fragment E). The average recovery in this experiment was 65%.

Lane 2 shows the position of unlabelled DNA fragments of known size (linear form of plasmid PCRI, PCRI DNA digested with restriction endonucleases EcoRI and XhoI, PCRI DNA digested with restriction endonuclease HaeIII; the numbers given are approximate and in base pairs, they have been derived partly by co-electrophoresis with ϕ X174 DNA (the sequence of which is known; Sanger et al., 1977) cut with restriction endonuclease Hae III). The fragments were located by staining with ethidium bromide and photography under UV-light (see section 2.2.4.5.). O: origin of electrophoresis (sample layering well); B: bromophenol blue dye marker.

products carrying a radioactive label were then determined by polyacrylamide gel electrophoresis followed by autoradiography; as size markers labelled DNA fragments of known length (viz. restriction fragments of plasmids PBR 322 and PCRI) were used.

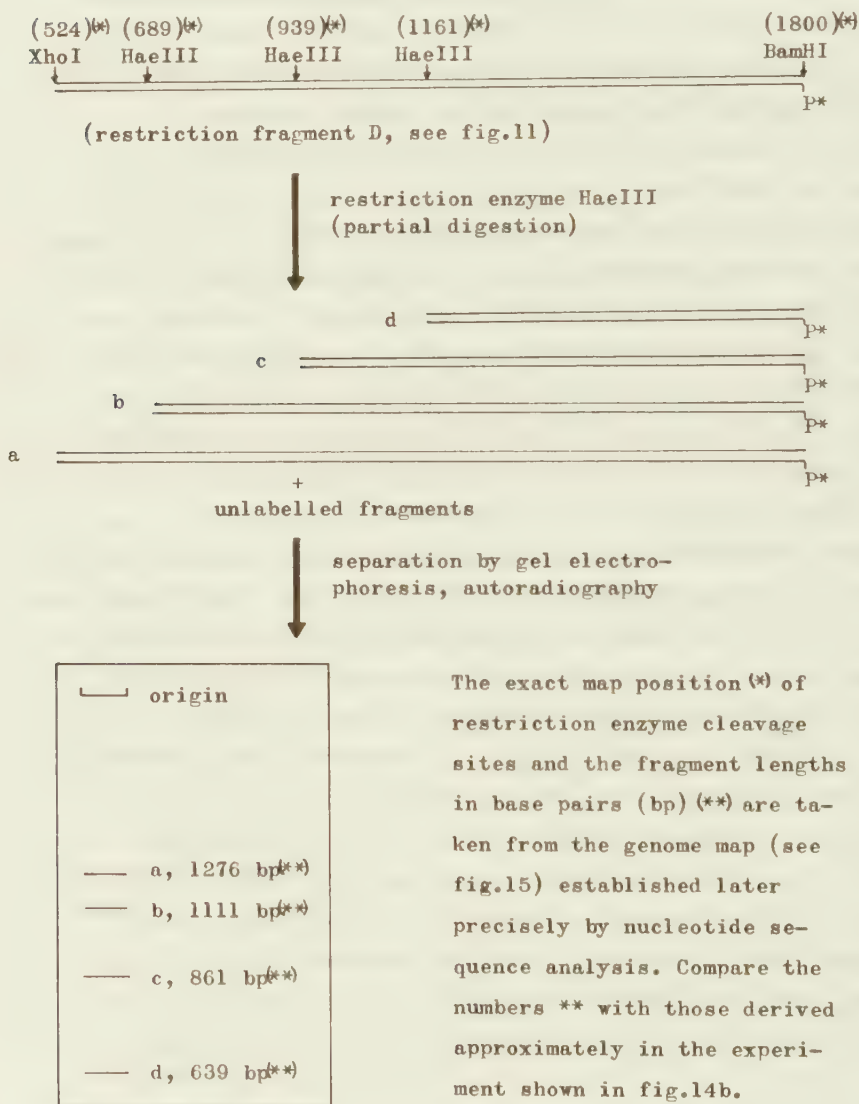
As schematically shown in fig.14a this technique in principle allows detection of every possible DNA fragment reaching from the original 5'-labelled end to one of the restriction cleavage sites. An actual example of this method (determination of HaeIII restriction sites on fragment D, see fig.11) is shown in fig.14b.

Maps for other restriction enzymes (AluI, HhaI, HinfI, HpaII, MboII and TaqI) were constructed in an analogous manner (gels not shown). In the case of restriction enzymes recognizing penta- or hexanucleotide sequences and hence cleaving only at a few sites of the $\Omega\beta$ sequence (AvaI, BamHI, BglII, BstEI, EcoRI, HaeII, PvuI, PvuII, RspI, SstI and XhoI) the digestions were carried out to completion and often combinations of these enzymes were used.

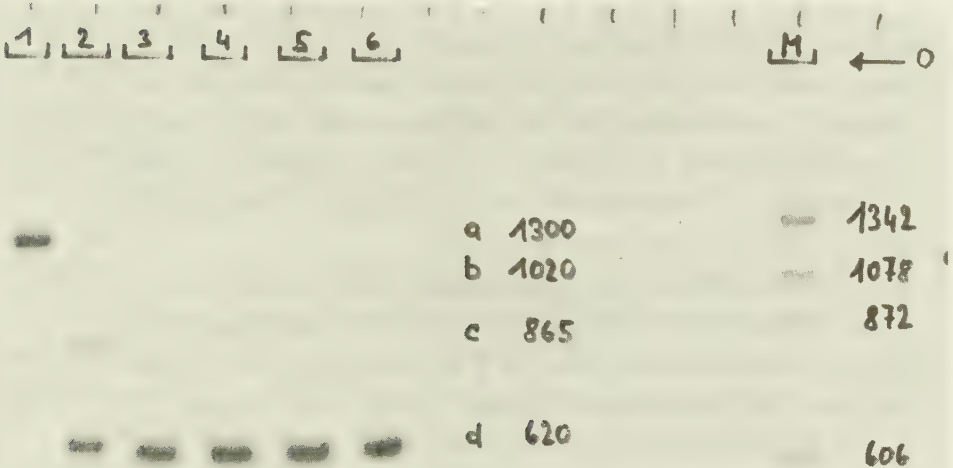
However, it must be mentioned that there are some drawbacks inherent in the mapping procedure involving partial digestion of end-labelled DNA fragments. In the following cases restriction sites are not detected. First, two sites may lie in close proximity; partial digestion will then yield two fragments of almost identical size which may not be separated by gel electrophoresis, suggesting only one restriction site. Second, a restriction site may be much more resistant to cleavage with a restriction enzyme than other restriction sites; this may either be due to (partial) modification of this site, or due to its particular surrounding nucleotide sequence. In such an instance, under the partial digestion conditions used, no cleavage at this particular site will occur (of course also classical

Figure 14: Restriction Mapping of Terminally Labelled DNA
Fragments Shown for the Mapping of Fragment D
(fig.11) with Endonuclease HaeIII

(a) Schematic Representation of the Procedure



(b) Polyacrylamide Gel Electrophoresis of the Fragments
Obtained by Partial Cleavage with Restriction Endo-
nuclease HaeIII



Restriction fragment D (see figs.11, 14a; 5 μ g containing 7400 Cherenkov cpm) in 35 μ l of Y-50 buffer (see 2.2.4.4.) was digested with restriction endonuclease BspRI (an isoschizomer of HaeIII; 0.5 U) at 37°C. After 0, 2, 5, 15, 35 and 90 min aliquots (5 μ l) were withdrawn, cooled to 0°C and EDTA/sucrose/bromophenol blue mixture was added (see section 2.2.4.4.). The samples were analyzed by polyacrylamide gel electrophoresis (6%; see 2.2.4.5. for experimental details). Lanes 1-6 in the autoradiograph show the digests after 0, 2, 5, 15, 35 and 90 min of incubation, respectively. Note that after 5 min of incubation (lane 3) the cleavage is essentially

complete, and that only in the digest corresponding to 2 min of incubation (lane 2) all possible labelled partial cleavage products are present. The approximate length in base-pairs of the digestion products a, b, c and d is indicated on the autoradiograph (compare these numbers with the exact numbers deduced from the known nucleotide sequence (see fig.14a): fragment a (experiment: 1300 bp, sequence: 1276 bp), fragment b (experiment 1020 bp, sequence: 1111 bp), fragment c (experiment: 865 bp, sequence: 861 bp) and fragment d (experiment: 620 bp, sequence: 639 bp)).

Lane M displays marker DNA fragments of known length labelled at the 5'-ends with [32 P]-phosphate (ϕ X174 DNA digested with restriction endonuclease HaeIII; 1 μ g containing 3500 Cherenkov cpm). Due to the known sequence of ϕ X174 DNA (Sanger et al., 1977) the exact length of the restriction fragments is known and indicated on the autoradiograph in numbers of base-pairs (bp).

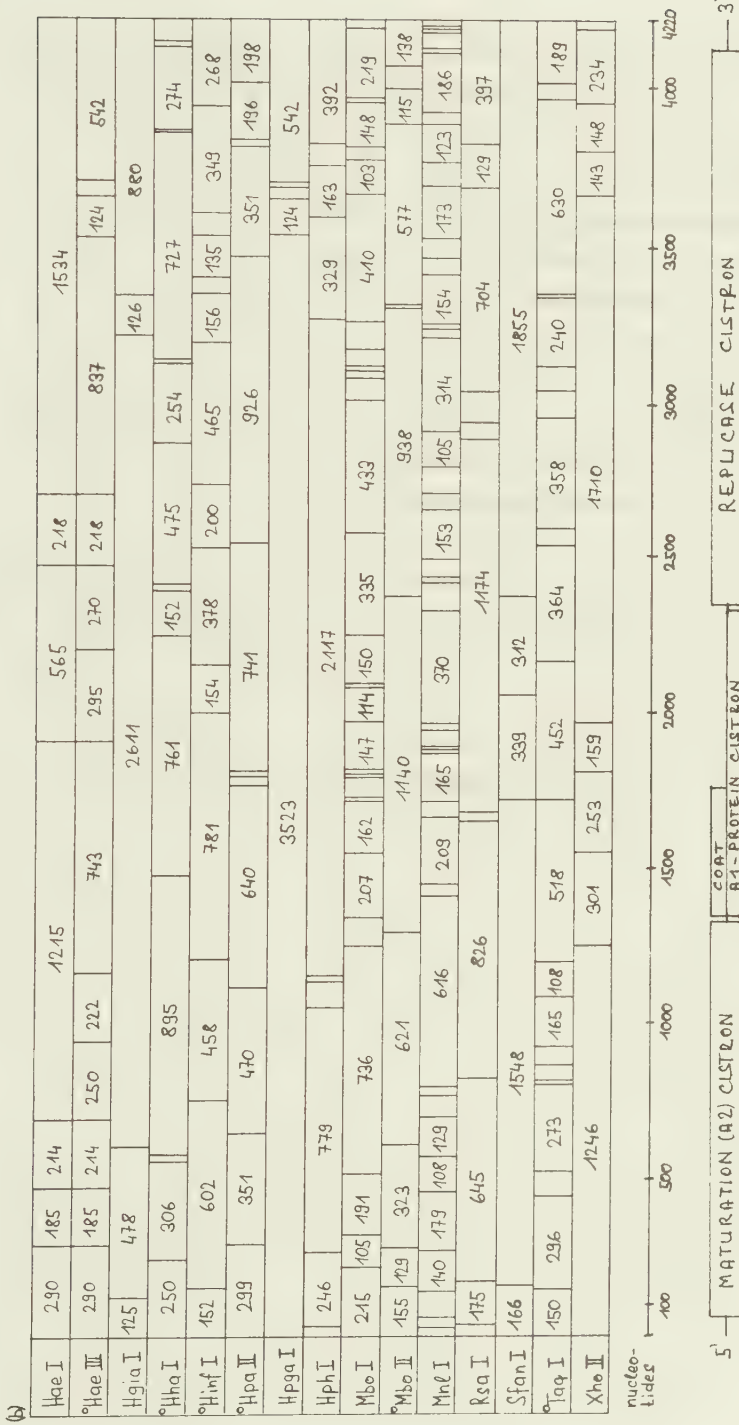
restriction mapping using complete digestion of the DNA does often not reveal fully modified sites). Difficulties of this type were found for the *TaqI* site at position 1712, which forms part of a methylated *MboI* site and was therefore not cleaved with restriction endonuclease *TaqI*. Another problem encountered are apparent restriction sites observed with the partial digestion method which are not actually present in the nucleotide sequence determined. The apparent site may be due either to a contaminating DNA fragment or to a contaminating restriction enzyme.

The restriction mapping method involving partial digestion yields the position of a restriction site (its distance from the labelled end, expressed in numbers of nucleotides) only with a precision of $\pm 10\%$. The deviations from the true values are due to some deviations from the linear relationship between ($1/\log$ molecular weight) and the electrophoretic mobility of the restriction fragments. Apparently the mobility of a DNA fragment in electrophoresis is dependent not only on its length (i.e. its molecular weight, expressed in numbers of base pairs), but to some degree also on its nucleotide composition.

Due to the pitfalls of the restriction mapping procedure involving partial digestion parallel confirmatory experiments using complete digestion of the DNA were necessary; fragments found after complete digestion with a particular restriction enzyme were then tentatively correlated with the restriction sites detected by the partial digestion method. As mentioned above, in some instances multiple digestions involving several different enzymes were used to obtain additional confirmation of the results.

The final restriction map of the DNA copy of the $\mu\beta$ genome is presented in fig.15.

Figure 15: Restriction Map of the $\Omega\beta$ Genome (continued)



The map shown was derived from the restriction mapping experiments using partial digestion of end-labelled fragments and complete digestion of unlabelled plasmid DNA containing the complete $\Omega\beta$ sequence described above. The data were supplemented and corrected with data from nucleotide sequence determination, where necessary. Map positions are in nucleotides, starting from the 5'-end of $\Omega\beta$ plus strand RNA.

Scheme (a) shows restriction sites for those enzymes cleaving the $\Omega\beta$ sequence once, twice or three times; scheme (b) gives the restriction sites for enzymes cleaving four or more times. The numbers in the boxes of scheme (b) indicate the length of the restriction fragments for a particular enzyme (only numbers larger than 100 nucleotides are given).

Restriction sites marked with open circles (e.g. $\circ$ BamHI or $\circ$ AluI) have been determined directly by experiment, all other sites have been deduced from the nucleotide sequence of the $\Omega\beta$ genome.

No restriction sites were detected for the following enzymes: BalI, EcoB, EcoK, HindIII, KpnI, PstI, SalI, SmaI and XbaI.

3.3. Nucleotide Sequence Determination in the $\alpha\beta$ Genome

The method for determination of nucleotide sequences in DNA introduced by Gilbert and Maxam in 1977 (Maxam & Gilbert, 1977) is now so widely used, and the principle, the technology and the occurrence and possible elimination of the numerous difficulties inherent in this method have been reviewed in such detail (Maxam & Gilbert, 1980), that it appears unnecessary to describe here again its theoretical and experimental features and the many problems which had to be gradually overcome in the earlier phases of this work. The principle of this sequencing approach (which under optimal conditions allows the direct reading of sequences of more than 200 nucleotides, starting from the terminal [^{32}P]-phosphate label of a DNA fragment) and the comparison with the methods available earlier for nucleotide sequence determination of RNA are briefly outlined in the Introduction (see section 1.5.).

To illustrate the method one autoradiogram of a Gilbert-Maxam sequencing gel is shown in fig. 16. In this experiment differentiation between the different base-specific chemical cleavage reactions (represented in the G-, purine (A+G), pyrimidine (T+C) and C-lanes) is good. The bands are narrow and easily distinguishable from each other, and generally the spaces between the bands are equal. However, in the lower part of the gel a compression (sequence ...CCGT..., indicated by a solid bracket) followed by a slight dilatation (sequence ...TAGT..., indicated by a dashed bracket) is visible. Such irregularities are due to strong secondary structures in the DNA fragments which even under the strong denaturing conditions used for the electrophoresis are not completely eliminated. Whereas in the example shown the reading is unambiguous sometimes two or more bands are superimposed, and thus one to a few nucleotides in a sequence may not be detected or the order may not be clear.

Figure 16: Gilbert-Maxam Nucleotide Sequence Determination between Nucleotides 1396 and 1442 of the $\alpha\beta$ Genome

G A T C



Plasmid Zp $\alpha\beta$ 32 DNA (100 μ g) was digested with a combination of endonucleases BamHI (100 U) and EcoRI (100 U) in a total volume of 525 μ l of Y-100 buffer (see 2.2.4.4.) for 60 min at 37°C. The cleavage products were resolved by preparative electrophoresis in an agarose gel (1.2%, see 2.2.4.5.). After superficial staining with ethidium bromide the DNA fragments were located under UV-light, and the 994 base pair fragment (reaching from the unique EcoRI site at position 806 in the $\alpha\beta$ genome to the BamHI site nearest to the 5'-end of the genome at position 1800, see fig.15) was isolated and purified on hydroxyapatite and DE-52 cellulose as described in section 2.2.4.5. (3.5 μ g of this restriction fragment were recovered, corresponding to 59% yield). 2 μ g of this preparation were digested with restriction endonuclease HhaI (5 U) in Y-150 buffer (25 μ l total volume) for 2 h at 37°C (see section 2.2.4.4.). The preparation was passed through Chelex-100, treated with bacterial alkaline phosphatase and afterwards with pronase/SDS/phenol (see section 2.2.5.), and the DNA was precipitated

1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2569
2570
2571
2572
2573
2574
2575
2576
2577
2578
2579
2580
2581
2582
2583
2584
2585
2586
2587
2588
2589
2590
2591
2592
2593
2594
2595
2596
2597
2598
2599
2600
2601
2602
2603
2604
2605
2606
2607
2608
2609
2610
2611
2612
2613
2614
2615
2616
2617
2618
2619
2620
2621
2622
2623
2624
2625
2626
2627
2628
2629
2630
2631
2632
2633
2634
2635
2636
2637
2638
2639
2640
2641
2642
2643
2644
2645
2646
2647
2648
2649
2650
2651
2652
2653
2654
2655
2656
2657
2658
2659
2660
2661
2662
2663
2664
2665
2666
2667
2668
2669
2670
2671
2672
2673
2674
2675
2676
2677
2678
2679
2680
2681
2682
2683
2684
2685
2686
2687
2688
2689
2690
2691
2692
2693
2694
2695
2696
2697
2698
2699
2700
2701
2702
2703
2704
2705
2706
2707
2708
2709
2710
2711
2712
2713
2714
2715
2716
2717
2718
2719
2720
2721
2722
2723
2724
2725
2726
2727
2728
2729
2730
2731
2732
2733
2734
2735
2736
2737
2738
2739
2740
2741
2742
2743
2744
2745
2746
2747
2748
2749
2750
2751
2752
2753
2754
2755
2756
2757
2758
2759
2760
2761
2762
2763
2764
2765
2766
2767
2768
2769
2770
2771
2772
2773
2774
2775
2776
2777
2778
2779
2780
2781
2782
2783
2784
2785
2786
2787
2788
2789
2790
2791
2792
2793
2794
2795
2796
2797
2798
2799
2800
2801
2802
2803
2804
2805
2806
2807
2808
2809
2810
2811
2812
2813
2814
2815
2816
2817
2818
2819
2820
2821
2822
2823
2824
2825
2826
2827
2828
2829
2830
2831
2832
2833
2834
2835
2836
2837
2838
2839
2840
2841
2842
2843
2844
2845
2846
2847
2848
2849
2850
2851
2852
2853
2854
2855
2856
2857
2858
2859
2860
2861
2862
2863
2864
2865
2866
2867
2868
2869
2870
2871
2872
2873
2874
2875
2876
2877
2878
2879
2880
2881
2882
2883
2884
2885
2886
2887
2888
2889
2890
2891
2892
2893
2894
2895
2896
2897
2898
2899
2900
2901
2902
2903
2904
2905
2906
2907
2908
2909
2910
2911
2912
2913
2914
2915
2916
2917
2918
2919
2920
2921
2922
2923
2924
2925
2926
2927
2928
2929
2930
2931
2932
2933
2934
2935
2936
2937
2938
2939
2940
2941
2942
2943
2944
2945
2946
2947
2948
2949
2950
2951
2952
2953
2954
2955
2956
2957
2958
2959
2960
2961
2962
2963
2964
2965
2966
2967
2968
2969
2970
2971
2972
2973
2974
2975
2976
2977
2978
2979
2980
2981
2982
2983
2984
2985
2986
2987
2988
2989
2990
2991
2992
2993
2994
2995
2996
2997
2998
2999
3000
3001
3002
3003
3004
3005
3006
3007
3008
3009
3010
3011
3012
3013
3014
3015
3016
3017
3018
3019
3020
3021
3022
3023
3024
3025
3026
3027
3028
3029
3030
3031
3032
3033
3034
3035
3036
3037
3038
3039
3040
3041
3042
3043
3044
3045
3046
3047
3048
3049
3050
3051
3052
3053
3054
3055
3056
3057
3058
3059
3060
3061
3062
3063
3064
3065
3066
3067
3068
3069
3070
3071
3072
3073
3074
3075
3076
3077
3078
3079
3080
3081
3082
3083
3084
3085
3086
3087
3088
3089
3090
3091
3092
3093
3094
3095
3096
3097
3098
3099
3100
3101
3102
3103
3104
3105
3106
3107
3108
3109
3110
3111
3112
3113
3114
3115
3116
3117
3118
3119
3120
3121
3122
3123
3124
3125
3126
3127
3128
3129
3130
3131
3132
3133
3134
3135
3136
3137
3138
3139
3140
3141
3142
3143
3144
3145
3146
3147
3148
3149
3150
3151
3152
3153
3154
3155
3156
3157
3158
3159
3160
3161
3162
3163
3164
3165
3166
3167
3168
3169
3170
3171
3172
3173
3174
3175
3176
3177
3178
3179
3180
3181
3182
3183
3184
3185
3186
3187
3188
3189
3190
3191
3192
3193
3194
3195
3196
3197
3198
3199
3200
3201
3202
3203
3204
3205
3206
3207
3208
3209
3210
3211
3212
3213
3214
3215
3216
3217
3218
3219
3220
3221
3222
3223
3224
3225
3226
3227
3228
3229
3230
3231
3232
3233
3234
3235
3236
3237
3238
3239
3240
3241
3242
3243
3244
3245
3246
3247
3248
3249
3250
3251
3252
3253
3254
3255
3256
3257
3258
3259
3260
3261
3262
3263
3264
3265
3266
3267
3268
3269
3270
3271
3272
3273
3274
3275
3276
3277
3278
3279
3280
3281
3282
3283
3284
3285
3286
3287
3288
3289
3290
3291
3292
3293
3294
3295
3296
3297
3298
3299
3300
3301
3302
3303
3304
3305
3306
3307
3308
3309
3310
3311
3312
3313
3314
3315
3316
3317
3318
3319
3320
3321
3322
3323
3324
3325
3326
3327
3328
3329
3330
3331
3332
3333
3334
3335
3336
3337
3338
3339
3340
3341
3342
3343
3344
3345
3346
3347
3348
3349
33

as described in section 2.2.1.1. The restriction fragments were then labelled by incubation with T4-polynucleotide kinase (0.5 U) and γ -[^{32}P]-ATP (15 pmol, specific activity 2×10^6 Cherenkov cpm/pmol) in a total volume of 15 μl at 37°C for 12 min (see section 2.2.5.). After chromatography through Sephadex G-100 (see section 2.2.1.3.) the DNA was precipitated (total yield was 4.8×10^6 Cherenkov cpm, 20% incorporation). The labelled DNA preparation was resolved into the component fragments by preparative electrophoresis in a polyacrylamide gel (7%, see section 2.2.4.5.), and the 668 base pair fragment (reaching from the EcoRI site at position 806 in the $\text{Q}\beta$ genome to the HhaI site at position 1474, see fig.15) was isolated and purified as described in section 2.2.4.5. (1.9×10^6 Cherenkov cpm were recovered after precipitation). To separate the two DNA strands the preparation was heat-denatured, the single strands were annealed to a 50-fold molar excess of $\text{Q}\beta$ plus strand RNA and the DNA plus and minus strands were isolated by chromatography on Sepharose CL-4B as described in section 2.2.5. (3.2×10^5 Cherenkov cpm were recovered in the minus strands carrying the 5'-terminal label at the HhaI cleavage site, 1.1×10^6 Cherenkov cpm were found in the plus strands carrying the label at the EcoRI cleavage site; see fig.17). The minus strand preparation was chemically degraded using the base-specific reactions according to Maxam and Gilbert (Maxam & Gilbert, 1977; Maxam & Gilbert, 1980; see also section 2.2.7.), the degradation products (10^4 Cherenkov cpm per base specific cleavage) were fractionated on a denaturing 20% polyacrylamide gel at 900 V for 12 h, and the gel was autoradiographed for two weeks (see section 2.2.7.). The nucleotide sequence was read off the autoradiogram from top to bottom as indicated, giving the sequence complementary to $\text{Q}\beta$ plus strand RNA between positions 1396 (GTTTGA... in the middle of the gel, (*) denotes position 1396) and 1442 (...TGCCGCA**), bottom, (**) indicates position 1442). Lane "G" shows G-specific, lane "A" purine-specific, lane "T" pyrimidine-specific and lane "C" C-specific cleavage products.

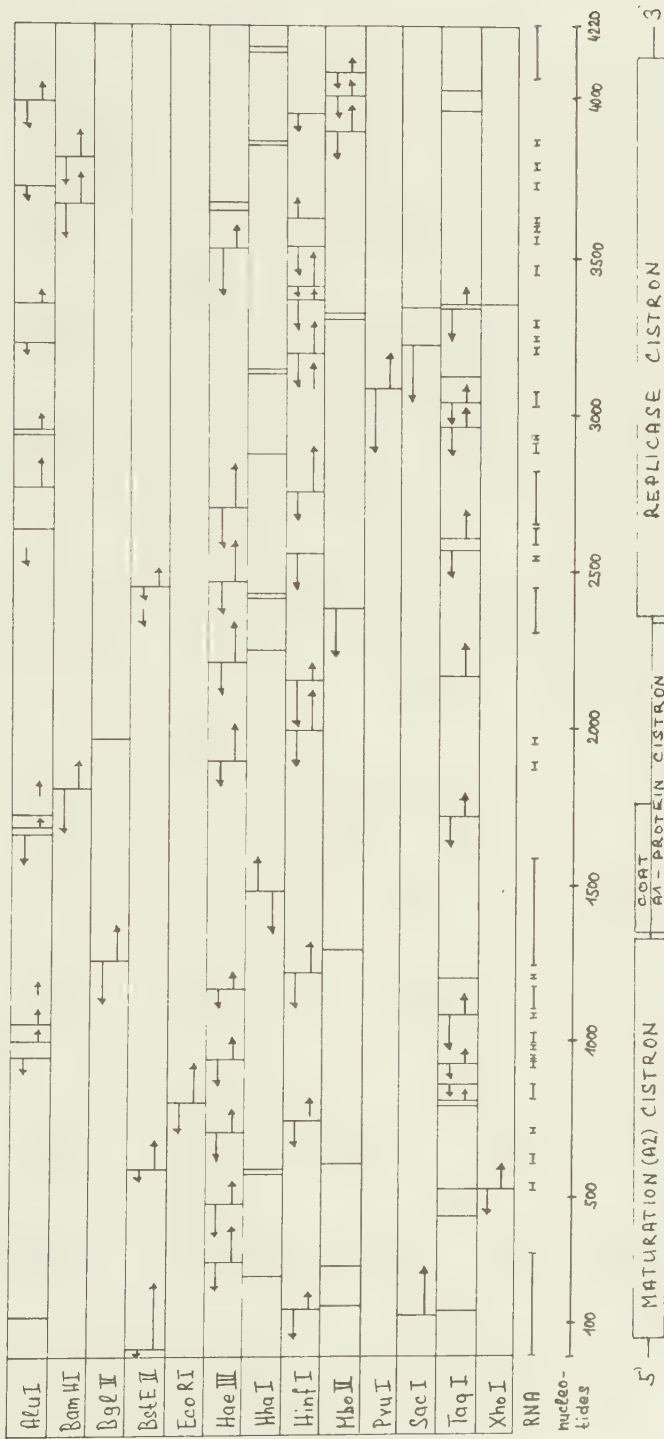
The strategy adopted for the sequence determination of the $\Omega\beta$ genome is outlined in fig.17: (i) Two independent and complementary sets of data were acquired for all sequences; either the sequence of both DNA strands was determined or, in regions where the RNA nucleotide or the protein aminoacid sequences were known, one of the DNA strands was analyzed. This proved to be necessary due to the artefacts mentioned above.

(ii) Sequence determination was carried out across every restriction enzyme cleavage site of the fragments selected for labelling with [32 P]-phosphate at the 5'-end. This was necessary to eliminate uncertainties which are frequently encountered in the reading of the first few nucleotides near the labelled end; it also ruled out the possibility that very short DNA regions between closely spaced restriction sites would go undetected.

To expedite the sequence analysis the $\Omega\beta$ genome was divided up between several workers. The sequence determined in this thesis comprises 24% of the total $\Omega\beta$ genome, and it spans the region from the unique EcoRI restriction site at position 806 to the BamHI restriction site nearest to the 5'-end at position 1800 (see fig.18). This region of 994 nucleotides contains approximately the last five hundred nucleotides of the A2-cistron, the region between the A2- and the coat cistrons, the entire coat cistron and finally another seventy nucleotides in the coat-A1 readthrough region. As shown in fig.17 restriction enzymes AluI, BamHI, BglII, EcoRI, HaeIII, HhaI, HinfI and TaqI were used for specific fragmentation of the $\Omega\beta$ DNA between positions 806 and 1800 and for generating sites for the introduction of 5'-terminal [32 P]-phosphate label into the DNA.

Whenever the results from the analysis of the two DNA strands or from one DNA strand and the RNA sequence data obtained pre-

Figure 17: Map of the Restriction Fragments Used for the Sequence Determination of the Q β Genome



The map given is a shortened version of the restriction map of the $\varrho\beta$ genome shown in fig.15. Map positions are in nucleotides, starting from the 5'-end of $\varrho\beta$ plus strand RNA. Only the restriction sites for enzymes used in the actual nucleotide sequence determination experiments are given in this figure.

The arrows in this map correspond to the sequences determined: each arrow represents an independent experiment, the readings starting at or near the 5'-terminal [^{32}P]-phosphate label at the respective restriction site. Arrows pointing to the left indicate sequence determination on the minus strand DNA; arrows pointing to the right show sequence determination on the plus strand DNA.

The RNA sequences determined prior to the nucleotide sequences on the DNA level (see section 1.5.) are shown by bars below the restriction map.

viously showed a discrepancy (this occurred in 84 out of 4220 (2%) of the nucleotide positions) there was no question which of the two versions to choose, since in all cases one set of data was unambiguous and the conflicting results could be explained. The nucleotide sequence of the entire Q β genome presented in fig.18 appears therefore free of errors as far as the experimental evidence is considered.

Possible sequence variations due to the retrotranscription of the RNA sequence into DNA and the heterogeneity inherent in the viral RNA preparation are dealt with in the Discussion (section 4.1.).

Figure 18: Nucleotide Sequence of the Q β Genome

(5) pppGGGGACCCCCUUUAGGGGGUCACCUCACACAGCAGUACUUCACUGAGUAUAAGAGGACA
 (START) 70 80 90 100 110 120
 UAUGCCUAAAUUACCGCGUGGUCUGCGUUUCGGAGCCGAUAUGAAAUUCUUAUGAUUU
 fM P K L P R G L R F G A D N E I L N D F
 130 140 150 160 170 180
 UCAGGAGCUCUGGUUCCAGACCUCUUUAUCGAAUCUCCGACACGCAUCCGUGGUACAC
 Q E L W F P D L F I E S S D T H P W Y T
 190 200 210 220 230 240
 ACUGAAGGGUCGUGUGUUGAACGCCACCUUGAUGAUCGUCUACCUAAUGUAGGGCGUG
 L K G R V L N A H L D D R L P N V G G R
 250 260 270 280 290 300
 CCAGGUAAGGCGCACUCCACAUCCGCGUCACCGUCCGAUUGCCUUCUACGGCCUUCGUCC
 Q V R R T P H R V T V P I A S S G L R P
 310 320 330 340 350 360
 GGUAACAACCGUUCAGUAUGAUCCCGCAGCACUAUCGUUCUUAUUGAACGCUCGUGUUGA
 V T T V Q Y D P A A L S F L L N A R V D
 370 380 390 400 410 420
 CUGGGAUUUCGGUAUAGGGCAUAGUGCGAACCUUGUCAUUAUGACUUCUGUUCGCAC
 W D F G N G D S A N L V I N D F L U R T
 430 440 450 460 470 480
 CUUUGCACCUAAGGAGUUUGAUUUUUGAACUCCUAGUCCUCGUUAUACUCAGGCCUU
 F A P K E F D F S N S L V P R Y T Q A F
 490 500 510 520 530 540
 CUCCGCGUUUAAUGCCAAGUAUGGCACUAUGAUCGGCGAAGGGCUCGAGACUAUAAAAUA
 S A F N A K Y G T M I G E G L E T I K Y
 550 560 570 580 590 600
 UCUCGGGCUUUUACUGCGAAGACUGCGUGAGGGUUAACGCGCUGUUAAGCGUGGGCAUUU
 L G L L L R R L R E G Y R A V K R G D L
 610 620 630 640 650 660
 ACGUGCUCUUCGUAGGGUUAUCCAGUCCUACCAUAAUGGUAAGUGGAAACGGGCUACUGC
 R A L R R V I Q S Y H N G K W K P A T A
 670 680 690 700 710 720
 UGGUAUUCUCUGGCUUGAAUUCGUUAUGGCCUUAUGCCUCUCUUUAUGACAUACAGAGA
 G N L W L E F R Y G L M P L F Y D I R D
 730 740 750 760 770 780
 UGUCAUUUAGACUGGCAGAACCUGCAUGAUAAAGAUCAACGCCUCCUUCGGUUUUCUGU
 V M L D W Q N R H D K I Q R L L R F S V
 790 800 810 820 830 840
 UGGUCACGGCGAGGAUUACGUUGUGCAAUUCGACAAUCUGUACCCUGCCGUUGCUUACU
 G H G E D Y V V E F D N L Y P A V A Y F

850	860	870	880	890	900
UAAACUGAAAGGGGAGAUUACACUCGAAACGCCGCAUCGUC AUGGCAUAUCUUACGCUAA					
K	L	K	G	E	I
910	920	930	940	950	960
CCGCGAAGGAUAUGCUGUUUUCGACAAACGGUCCCUUCGGCCUGUGUCCGAUUGGAAGGA					
R	E	G	Y	A	V
970	980	990	1000	1010	1020
GCUUGCCACUGCAUUCAUCAAUCCGCAUGAAGUUGCUUGGGAGUUAACUCCCUACAGCUU					
L	A	T	A	F	I
1030	1040	1050	1060	1070	1080
CGUUGUUGAUUGGUUCUUGAAUGUUGGUGACAUACUUGCUCACAAAGGUCAGCUAUUAUCA					
V	V	D	W	F	L
1090	1100	1110	1120	1130	1140
UAAUAUCGAUAUUGUAGACGGCUUUGACAGACGUGACAUCCGGCUCAAAUCUUUCACCAU					
N	I	D	I	V	D
1150	1160	1170	1180	1190	1200
AAAAGGUGAACGAAAUUGGGCGGCCUGUUAACGUUUCGCUAGCCUGUCUGCUGCUGCAUUU					
K	G	E	R	N	G
1210	1220	1230	1240	1250	1260
AUUUUACAGCCGACUCCAUACGAGCAUCUUCCGUUCGCUACACUAGAUUCUUGAUACUAC					
F	Y	S	R	L	H
1270	1280	1290	1300	1310	1320
CUUUAGUUCGUUUAACACGUCUUGAUAGUAUCUUUUUAUUAACCCAACCGGUAAGCGG					
F	S	S	F	K	H
(END OF)	1330	1340 (START OF)	1350	1360	1370
UUGAA	ACUUUGGGUCAAUUUGAUC	AUGGCAAAA	AUAGAGACUGU	ACUUUAGGUAAC	AUC
		fM	A	K	L
1390	1400	1410	1420	1430	1440
GGGAAAGAUAGGAAAACAAACUCUGGUCCUCAAUCCGCGUGGGGUAAAUCCACUAACGGC					
G	K	D	G	K	Q
1450	1460	1470	1480	1490	1500
GUUGCCUCGCUUACAAAGCGGGUGCAGUUCUGCGCUGGAGAAGCGUGUUACCGUUUCG					
V	A	S	L	S	Q
1510	1520	1530	1540	1550	1560
GUAUCUCAGCCUUCUCGCAAUUCGUAAGAACUACAAGGUCCAGGUUAAGAUCCAGAACCCG					
V	S	Q	P	S	R
1570	1580	1590	1600	1610	1620
ACCGCUUGCACUGCAAAACGGUUCUUGUGACCCAUCCGUUACUCGCCAAGCAUAUGCUGAC					
T	A	C	T	A	N
1630	1640	1650	1660	1670	1680
GUGACCUUUCGUUCACGCAGUAUAGUACCGAUGAGGAACGAGCUUUUGUUCGUACAGAG					
V	T	F	S	F	T
1690	1700	1710	1720	1730	1740
CUUGCUGCUCUGCUCGCUAGUCCUCUGCUGAUCGAUGCUAUUGAUCAGCUGAACCCAGCG					
L	A	A	L	L	A

(END COAT;
(READTHROUGH AM)) 1750 1760 1770 1780 1790 1800
UAUUGAACACUGCUCUAUUGCCGGUGGUGGCUCAGGUGUAAAACCCGAUCCGGUUAUUCGG
Y (W) T L L I A G G G S G S K P D P V I P
1810 1820 1830 1840 1850 1860
GAUCCACCGAUUGAUCCGCCGCCAGGGACAGGUAAGUAUACCUGUCCCUUCGCAAUUUGG
D P P I D P P P G T G K Y T C P F A I W
1870 1880 1890 1900 1910 1920
UCCCUAGAGGAGGUUUACGAGCCUCCUACUAAGAACCGACCGUGGGCUAUCUAUAUUGCU
S L E E V Y E P P T K N R P W P I Y N A
1930 1940 1950 1960 1970 1980
GUUGAACUCCAGCCUCGGCAAUUUGAUGUUGCCCUCAAAGAUUUUUGGGCAAUACAAAG
V E L Q P R E F D V A L K D L L G N T K
1990 2000 2010 2020 2030 2040
UGGCGUGAUUGGGAUUCUCGGCUUAGUUAUACCACGUUCCGCGGUUGCCGUGGCAAUUGGU
W R D W D S R L S Y T T F R G C R G N G
2050 2060 2070 2080 2090 2100
UAUAUUGACCUUGAUUGCACUUAUCUUGCUACUGAUCAGGCUAUGCGUGAUCAGAAGUAU
Y I D L D A T Y L A T D Q A N R D Q K Y
2110 2120 2130 2140 2150 2160
GAUAUUCGCGAAAGGCAAAGAAACCGUGUCUUUCGGUAACAUUGAGCGAUUCAUUUAUCUU
D I R E G K K P G A F G N I E R F I Y L
2170 2180 2190 2200 2210 2220
AAGUCGAUAAAUGCUUAUUGCUCUCUUAAGCGAUUUGCGGCCUAUCACGCCGAUGGCGUG
K S I N A Y C S L S D I A A Y H A D G V
2230 2240 2250 2260 2270 2280
AUAGUUGGCUUUUGGCGCGAUCCAUCCAGUGGUGGUGGCCAUACCGUUUGACUUCACUAAG
I V G F W R D P S S G G A I P F D F T K
2290 2300 2310 2320 2330 (END AM) 2340
UUUGAUAAAGACUAAAUGUCCUAUUCAAAGCCGUGAUAGUCGUUCCUCGUGCUUAGUAACUA
F D K T K C P I Q A V I V V P R A
2350 (START
(REPLICA)) 2360 2370 2380 2390 2400
AGGAUGAAAUGCAUUCUAAGACAGCAUCUUCGCGUAACUCUCUCAGCGCACAAUUGCGC
fm S K T A S S R N S L S A Q L R
2410 2420 2430 2440 2450 2460
CGAGCCCGCAACACAAGAAUUGAGGUUGAAGGUAACCUCGCACUUUCCAUUGCCAACGAU
R A A N T R I E V E G N L A L S I A N D
2470 2480 2490 2500 2510 2520
UUACUGUUGGCCUAUGGUCAGUCGCCAUUUAACUCUGAGGCUGAGUGUAUUUCAUUCAGC
L L L A Y Y Q S P F N S E A E C I S F S
2530 2540 2550 2560 2570 2580
CCGGAUUCGACGGGACCCCGGAUGACUUUAGGAUAAAUAUCUUAAGCCGAGAUCAUG
P R F D G T P D D F R I N Y L K A E I M
2590 2600 2610 2620 2630 2640
UCGAAGUAUGACGACUUCAGCCUAGGUUAUGAUACCGAAGCUGUUGCCUGGGAGAGAAGUUC
S K Y D D F S L G I D T E A V A W E K F

2650	2660	2670	2680	2690	2700
CUGGCAGCAGAGGCUGAAUGUGCUUUAACGAACGCUCGUCUCUAUAGGCCUGACUACAGU					
L A A E A E C A L T N A R L Y R P D Y S					
2710	2720	2730	2740	2750	2760
GAGGAUUUCAAUUUCUCACUGGGCGAGUCAUGUAUACACAUGGCCUGUAGAAAAAUAGCC					
E D F N F S L G E S C I H M A R R K I A					
2770	2780	2790	2800	2810	2820
AAGCUAAUAGGAGAUGUCCGUGCGUUGAGGGUAUGUUGCGUCACUGCCGAUUUUCUGGC					
K L I G D V P S V E G M L R H C R F S G					
2830	2840	2850	2860	2870	2880
GGUGCUACAACAACGAUAUACCGUUCGUACGGUCAUCCGUCUUAAGUUUGCGCUUCCG					
G A T T T N N R S Y G H P S F K F A L P					
2890	2900	2910	2920	2930	2940
CAAGCGUGUACGCCUCGGGCUUUGAAGUAUGUUUAGCUCUCAGAGCUUCUACACAUUUC					
Q A C T P R A L K Y V L A L R A S T H F					
2950	2960	2970	2980	2990	3000
GAUACCAGAAUUUCUGAUUUAGCCUUUUAAUAAAGCAGUUACUGUACCUAAGAACAGU					
D T R I S D I S P F N K A V T V P K N S					
3010	3020	3030	3040	3050	3060
AAGACAGAUUCGUUGUAUUGCUAUCGAACCGGUUGGAUAUGUUUUUCCAACUGGGUAUC					
K T D R C I A I E P G W N M F F Q L G I					
3070	3080	3090	3100	3110	3120
GGUGGCAUUCUACGCGAUCGGUUGCGUUGCUGGGGUAUCGAUCUGAAUGAUCAGACGAUA					
G G I L R D R L R C W G I D L N D Q T I					
3130	3140	3150	3160	3170	3180
AAUCAGCGCCGCGCUCACGAAGGCUCGUAACUAAUAAACUAGCAACCGGUUGAUCUCUCA					
N Q R R A H E G S V T N N L A T V D L S					
3190	3200	3210	3220	3230	3240
GCGGCAAGCGAUUCUAUAUCUCUUGCCUCUGUGAGCUCUUAUUGCCCUAGGCUGGUU					
A A S D S I S L A L C E L L L P L G W F					
3250	3260	3270	3280	3290	3300
GAGGUUCUUAUGGACCUCAGAUACCUAAGGGGCGAUUGCCUGACGGUAGUGUUGUACC					
E V L M D L R S P K G R L P D G S V V T					
3310	3320	3330	3340	3350	3360
UACGAGAAGAUUUCUUCUAUGGGUAACGGUUAACAUUCGAGCUCGAGUCGCUUAAUUUU					
Y E K I S S M G N G Y T F E L E S L I F					
3370	3380	3390	3400	3410	3420
GCUUCUCUCGCUUGUCCGUUGUGAGAUACUGGACUUAAGACUCGUCUGAGGUCACUGUU					
A S L A R S V C E I L D L D S S E V T V					
3430	3440	3450	3460	3470	3480
UACGGAGACGAUAUAUUUUUACCGUCCUGUGCAGUCCUGCCUCCGGAAGUUUUUAG					
Y G D D I I L P S C A V P A L R E V F K					
3490	3500	3510	3520	3530	3540
UAUGUUGGUUUUACGACCAAUACUAAAAAGACUUUUUCCGAGGGGCCGUUCAGAGAGUCG					
Y V G F T T N T K K T F S E G P F R E S					

3550	3560	3570	3580	3590	3600
UGCGGCAAGCACUACUAAUUCUGGCGUAGAUGUUAACUCCCUUUAACAUAACGUCACCGUAUA					
C G K H Y Y S G V D V T P F Y I R H R I					
3610	3620	3630	3640	3650	3660
GUGAGUCCUGCCGAUUUAAUACUGGUUUUGAAUAACCUAAUACGUGGUGGCCACAAUUGAC					
V S P A D L I L V L N N L Y R W A T I D					
3670	3680	3690	3700	3710	3720
GGCGUAUGGGGAUCCUAGGGCCCAUUCUGUGUACCUCAAGUAUCGUAAGUUGCCGCCUAAA					
G V W D P R A H S V Y L K Y R K L P P K					
3730	3740	3750	3760	3770	3780
CAGCUGCAACGUAAUACUUAUACUGAUGGUUACGGUGAUGGUGCCCUUGCGUGGAUCGGUC					
Q L Q R N T I P D G Y G D G A L V G S V					
3790	3800	3810	3820	3830	3840
CUAAUAUCCUUUCGCGAAAAACCGCGGGUGGAUCCGGUACGUACCGGUGAUUACGGAC					
L I N P F A K N R G W I R Y V P V I T D					
3850	3860	3870	3880	3890	3900
CAUACAAGGGACCGAGAGCGCGCUGAGUUGGGGUCGUAUCUCUACGACCUCUUCUGCGU					
H T R D R E R A E L G S Y L Y D L F S R					
3910	3920	3930	3940	3950	3960
UGUCUCUGGGAAGUAACGAUGGGUUGCCUCUUAAGGGGUCCAUCCGGGUGCGAUUCUGCG					
C L S E S N D G L P L R G P S G C D S A					
3970	3980	3990	4000	4010	4020
GAUCUAUUUGCCAUCGAUCAGCUUAUCUGUAGGAGUAAUCCUACGAAGAUAAAGCAGGUCC					
D L F A I D Q L I C R S N P T K I S R S					
4030	4040	4050	4060	4070	4080
ACCGGC AAAUUCGAUAUACAGUAUAUCGCGUGCAGUAGCCGUGUUCUGGCACCCUACGGG					
S G K F D I Q Y I A C S S R V I A P Y G					
4090	4100	4110	4120	4130	4140
GUCUUCACGGGCACGAAGGUUGCGUCUCUACACGAGGCGUAAACUGGGGGAGGGCGCCAA					
V F Q G T K V A S L H E A					
4150	4160	4170	4180	4190	4200
UAUGGCGCCUAAUUGUGAAUAAAUUAUCACAAUUAUCUCUUAACGAGUGAGAGGGGAUCUG					
4210	4220				
CUUUGCCUCUCUCCUCCCA _{OH} (3')					

The sequence shown is the one of $Q\beta$ plus strand RNA; the numbering of the nucleotides is from the 5'- to the 3'-end of the RNA (the last digit of each number indicates the position). The aminoacid sequences of all $Q\beta$ specific gene products (as derived from the nucleotide sequence) are shown below the sequence of the RNA, using the one-letter code for aminoacids (Dayhoff, 1972).

4. DISCUSSION

4.1. Fidelity of Q β RNA Retrotranscription

One of the most important questions arising, when transcribing an RNA into DNA, is whether the DNA transcript reflects the exact nucleotide sequence of the RNA template.

It is known that AMV reverse transcriptase, used for the transcription of Q β RNA, has a relatively high error rate, about 10^{-4} /nucleotide (Sirover et al., 1977; Maniatis et al., 1976). Thus transcription of the Q β genome by this enzyme might introduce about one nucleotide change per genome length.

Another source of sequence errors could conceivably occur during the treatment of the RNA-DNA hybrid by strong alkali to eliminate the RNA strand. At the high pH used deamination of deoxycytidine to deoxyuridine may occur in the DNA strand (Sirover et al., 1977). During second transcription of this DNA strand A-residues would be incorporated at these sites. Thus, in principle, G-residues in the RNA sequence could be replaced by A-residues in the DNA plus strand sequence.

However, as mentioned in the Introduction (section 1.4.), plasmids containing Q β full-size sequences give rise to viable phage in bacteria (Taniguchi et al., 1978). These DNA sequences represent either the "wild type" RNA sequence (see below) or contain only silent mutations. All sequence determinations were carried out on such plasmids: therefore the derived sequence is that of a biologically active phage. Furthermore, since three out of four plasmids containing the entire Q β sequence gave rise to phage in E.coli (Taniguchi et al., 1978) it is clear that the RNA transcription and the cloning of the transcripts introduce very few sequence errors.

It should be noted here that Q β RNA itself contains a certain amount of variability in its nucleotide sequence. Since Q β replicase replicating the Q β genome introduces about one chan-

ge per 10^4 nucleotides (Domingo et al., 1976; Batschelet & Weissmann, 1976; Domingo et al., 1978), and since at least some of these changes leave the biological competence unaltered (Domingo et al., 1976), it may be erroneous to define a unique nucleotide sequence for wildtype $\varrho\beta$ RNA.

Therefore $\varrho\beta$ RNA from uncloned phage (or even $\varrho\beta$ RNA from cloned phage which has undergone many replicative cycles before the RNA could be prepared) will consist of a mixture of RNAs of slightly variant sequences. Some of these RNAs will show biological competence, while others will be non-viable in vivo. Conventional sequence determination on the RNA level (see section 1.4.) will present an averaged-out nucleotide sequence. The use of DNA cloning techniques, however, will select one of the several biologically relevant RNA sequences. The selected sequence found in the DNA might not necessarily represent the "average" sequence most commonly found in a phage stock.

Initially sequence determination was carried out on plasmids containing only parts of the $\varrho\beta$ sequence (Mekler et al., 1977; see section 3.1.). These had been constructed from uncloned phage RNA and could obviously not be checked for biological competence. But nucleotide sequence determination of several $\varrho\beta$ plasmids derived from uncloned RNA should in principle reveal sites of preferential mutation. We have therefore compared various sequences derived from such plasmids (a total of approximately 300 nucleotides from two regions corresponding to positions 450 to 600 in the $\varrho\beta$ genome (in the vicinity of the 5'-terminal XhoI restriction endonuclease cleavage site) and positions 700 to 850 (around the unique EcoRI restriction site). Only one sequence variant was found (at position 791): plasmid clones P $\varrho\beta$ 160 (partial $\varrho\beta$ sequence) and Zp $\varrho\beta$ 32 (complete $\varrho\beta$ sequence; Taniguchi et al., 1978) show

a G-residue at this position, while clone PQ β 61 (partial Q β sequence) shows an A-residue at position 791. Since out of 300 nucleotides of sequences compared only one variant was found neither of the regions analyzed seems to be a region of especially frequent mutational nucleotide changes. However, many more comparisons of independent clones would be necessary to define regions of high or low incidence of mutation.

4.2. Relationships between Bacteriophages Q β and MS2

4.2.1. Introduction

The RNA phages Q β and MS2 have been subject to scrutiny since the 1960's. Although they differ in serotype, genome length and overall nucleotide sequence (no cross-hybridization of their respective RNA plus- and minus strands was detectable; Weissmann & Ochoa, 1967) common features have been found ever since their discovery. Evolutionary relationships between the phage groups I and III (with MS2 and Q β as their most representative members; see section 1.1.) have been proposed (Konigsberg et al., 1970; Hofstetter et al., 1974) on the following grounds: From a structural point of view both MS2 and Q β phages are similar in their size and symmetry, they are built up from similar proteins encapsidating the phage RNA and their respective major coat proteins show some slight homology in their aminoacid sequence (Weber & Konigsberg, 1975). From a functional point of view both phage genomes fulfil a dual role: on one side they serve to store and replicate the genetic information of the viruses, on the other they function directly as polycistronic messengers for the production of three major gene products (coat-, maturation-, replicase polypeptides). Furthermore, the arrangement of the respective cistrons is the same, and many of their physiological features

(involvement of the same host proteins for virus replication, regulation of the gene expression during the lytic cycle) are also similar. A concise summary of all these aspects is given in sections 1.1. and 1.2.

Elucidation of the complete nucleotide sequence of MS2 RNA (Nichols, 1970; De Wachter et al., 1971; Contreras et al., 1972; Min Jou et al., 1972; Rensing et al., 1973; Fiers et al., 1975; Fiers, 1975; Vandenberghe et al., 1975; Fiers et al., 1976) and of $\varrho\beta$ RNA (this thesis and paper in preparation) allows the comparison of the two phages on a more detailed level.

Anticipating the results discussed in the following sections (primarily considering the aminoacid sequence homologies of the replicase β -subunits of the two phages; see section 4.2.2.2.) it appears now beyond doubt that both phages are descendants from a common ancestor and that the structural and functional similarities are not due to a convergent evolution of initially different phages.

The first level of comparison between $\varrho\beta$ and MS2 phages concerns the primary structure of the two phage genomes; the second level would be concerned with the structures resulting from the folding of the RNAs (secondary and higher order structures). Many questions regarding the layout of the genetic information (precise definition of major and minor cistrons, codon usage), the properties of the gene products and the evolution of the phages can be answered by consideration of the primary structures, as will be shown in the following sections; the understanding of many functional aspects, especially the properties of the regions involved in recognition and regulation, would require knowledge of higher order structure.

In the case of phage MS2, Fiers and coworkers have published

their "flower model" for the complete secondary structure of the genomic RNA (Min Jou et al., 1972; Fiers et al., 1975; Fiers et al., 1976). The proposed structures have emerged to a large extent from the type of sequence analysis involving RNA fragments produced by mild treatment of MS2 RNA with ribonuclease (see section 1.5.): under conditions of partial RNase digestion nucleolytic cleavages are introduced predominantly into single stranded regions and into loops and bulges rather than into the stem regions of the proposed hairpins (Gould, 1967; Min Jou et al., 1968; De Wachter et al., 1971). Moreover, some RNA segments thought to constitute the stems of hairpin-loop structures remain hydrogen-bonded and migrate as complexes during gel electrophoresis under non-denaturing conditions, although they are no longer covalently connected after nuclease treatment; only upon electrophoresis under denaturing conditions they separate and their sequence can be determined (Jeppesen et al., 1970; Nichols, 1970; Cory et al., 1972; Min Jou et al., 1972).

For $\Omega\beta$ such partial nuclease digestion data is very scant since most sequence determinations were carried out on short RNA segments synthesized in vitro (Billeter et al., 1969a; Billeter et al., 1969b; Kolakofsky & Weissmann, 1971a; Kolakofsky & Weissmann, 1971b) and on the DNA rather than the RNA (this work; Mekler et al., 1978). Therefore the attempt to formulate secondary structures for $\Omega\beta$ RNA relies entirely on the inspection of base-pairing matrix schemes (Tinoco et al., 1971), which was greatly facilitated by a computer program yielding all possible base-pairings for a given nucleotide sequence (computer program by courtesy of Dr. R. Portmann, Universität Zürich). Selection among the possible structures was made for the one with the highest thermodynamic stability according to the published rules (Delisi & Crothers, 1971; Tinoco et al.,

1971; Gralla & Crothers, 1973). Due to a limitation in the computer program possible secondary structures of only comparatively short regions (up to 120 nucleotides) could be investigated with ease. With the present data it is therefore not possible to deduce a definitive and detailed secondary (and higher order) structure for the whole of $\varphi\beta$ RNA, and even secondary structures for short regions, as shown in the course of this discussion for regions of particular interest, must be considered as very tentative.

4.2.2. Comparison of Coding Sequences

The overall organization of the genomes of MS2 and $\varphi\beta$ has already been shown in fig.1, and the products of the major and minor cistrons were reviewed in section 1.2. Here the exact locations, lengths and nucleotide compositions of the various genome regions are compared (table 4).

4.2.2.1. Reading Frame and Codon Usage

The similarity between phages $\varphi\beta$ and MS2, which is already reflected in the overall organization of their genomes, becomes more evident when the layout of the genetic information is compared. For this reason the reading frames and the codons used in the frames will be compared in this section, using the known nucleotide sequence of the two phage RNAs. This finally permits the comparison of phage proteins, as will be shown in section 4.2.2.2.

Determination of reading frame: Since genetic information in the RNA is laid down in a comma-free triplet code, the first step in deducing an aminoacid sequence from a nucleotide sequence is to find the "sense-carrying" reading frame. This is quite obvious in any nucleotide sequence of appreciable length since in the other two frames frequent termination

Table 4: Length and Composition of $\varphi\beta$ and MS2 Cistrons and Non-Translated Regions

	$\varphi\beta$					MS2				
	length, nucleo- tides	nucleotide composition (%)				length, nucleo- tides	nucleotide composition (%)			
		A	G	C	U		A	G	C	U
5'-extracis- tronic region	61	26.2	23.0	27.9	22.9	129	17.8	25.6	26.4	30.2
maturation cistron ($\varphi\beta$: A2, MS2: A)	1263	22.2	22.3	24.7	30.8	1182	23.6	25.9	26.3	24.2
maturation- coat inter- cistronic region	20	25.0	20.0	15.0	40.0	23	26.1	26.1	30.4	17.4
coat cistron	402	23.9	23.9	26.1	26.1	393	26.2	22.9	26.5	24.4
Al-cistron	990	22.8	24.4	24.8	28.0	(not present)				
coat/Al- replicase intercistro- nic region	18	44.4	22.2	11.1	22.3	33	30.3	21.2	30.3	18.3
replicase cistron	1770	22.8	25.0	23.6	28.2	1638	23.2	25.8	25.6	25.4
3'-extracis- tronic region	98	22.4	25.5	26.6	25.5	171	20.5	32.7	16.4	30.4
total genome	4220	22.8	24.2	23.9	29.1	3569	23.4	25.8	26.3	24.5

The $\varphi\beta$ genome is 651 nucleotides longer than that of MS2. The difference is mainly due to the "readthrough" region in $\varphi\beta$. Furthermore both the maturation (A2) cistron and the replicase cistrons are longer in $\varphi\beta$ (plus 27 and 132 nucleotides, respectively); all non-translated regions are longer in MS2.

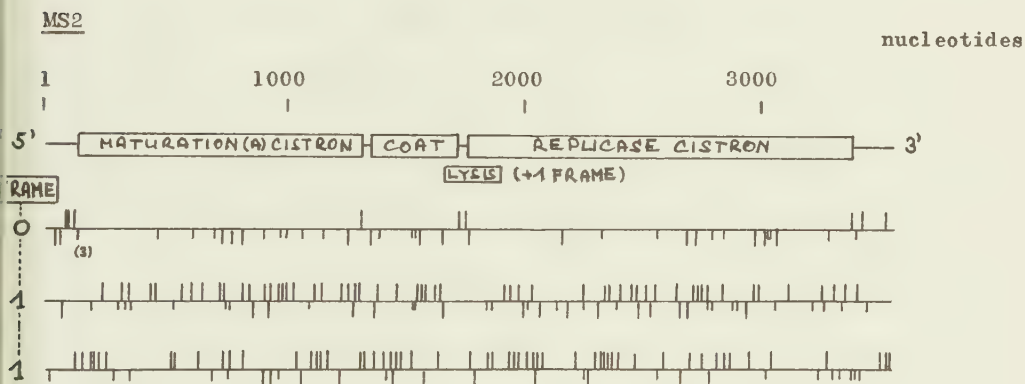
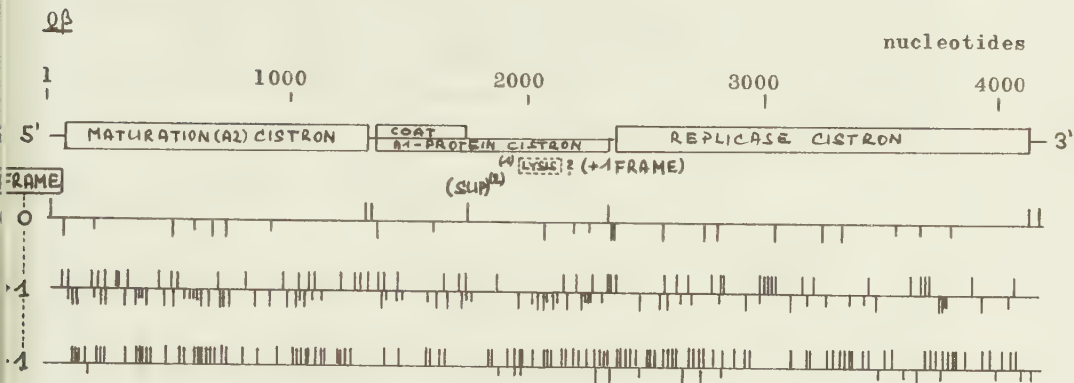
Both phages have a similar base composition: $\varphi\beta$ RNA is more rich in U, MS2 RNA more rich in C; both RNAs contain less A than on a random basis. The overall base composition of both phage RNAs is faithfully reflected in the base composition of each individual cistron.

codons (statistically one termination codon per 20.3 aminoacid codons) prohibit extensive translation of the RNA. At the same time the distribution of the initiation codons (generally AUG; GUG in the case of the MS2 maturation protein) within the three reading frames defines possible starting points for translation. However, to allow actual initiation of translation at these sites additional constraints regarding the nucleotide sequence preceding the "proper" initiation codon must exist (Shine & Dalgarno, 1974; Steitz & Jakes, 1975; Taniguchi & Weissmann, 1978; see also section 4.2.3.3.).

The distribution of initiation and termination codons in every theoretically possible reading frame of $\Omega\beta$ and MS2 RNA is presented in fig.19. It is clear that the reading frames for the major gene products (the zero frames) emerge as the only ones devoid of termination codons (with the exception of the terminators defining the end of a cistron). In $\Omega\beta$, in the +1 frame, the average distance separating two termination codons is 77 nucleotides, while in the -1 frame it is 36 nucleotides (on a random basis one terminator per 64 nucleotides is expected). In MS2 the distribution of the termination codons in the two non-zero frames is more regular: in both the +1 and -1 frames the average distance between two terminators is 61 nucleotides. Exceptions to this clearcut distinction between coding- and noncoding frames are discussed in section 4.2.4.2.

The distribution of the AUG initiation codons (GUG codons are disregarded in this context, since this codon is only used once as an initiator in the case of the MS2 A-cistron) is very unequal in $\Omega\beta$: in the zero reading frame an average distance of 300 nucleotides between two initiators is found (on a random basis one AUG per 192 nucleotides is expected). In the +1 frame the average distance is 105 nucleotides, while in the -1 frame it is 845 nucleotides. In MS2 the distribution of the AUG initiation codons in all three reading frames is roughly

Figure 19: Distribution of Initiation and Termination Codons
in the Genomes of Q β and MS2



Marks below the lines indicate AUG codons (full-length dash) or GUG codons (half-length dash); marks above the lines indicate the termination codons UAA, UAG and UGA.

- (1) possible lysis polypeptide in $\varrho\beta$ (see section 4.2.4.2.)
- (2) the UGA termination codon of $\varrho\beta$ coat cistron (position 1744) is suppressed for readthrough-translation of A1-cistron (see 1.2.2.)
- (3) initiation codon of MS2 maturation cistron is GUG

(continued on next page)

- (4) note that there is a change of frames between cistrons: the reading frames are therefore defined for the cistrons proper including the adjacent intercistronic regions up to the beginning of the next cistron. The frames in the 5'-extracistronic regions are defined with respect to the maturation cistron; the frame in the corresponding 3'-terminal regions are defined with respect to the frame of the replicase cistrons

equal: the average distance between two AUG codons is found to be 255 nucleotides.

The highly unequal distribution of initiation and termination codons in the +1 and -1 reading frames of $\varrho\beta$ may not be fortuitous. The distribution pattern of initiation and termination codons in $\varrho\beta$ and MS2 is very different and does not reveal any relationships between the two phages.

Additional minor gene products (see section 4.2.4.2. for a possible lysis polypeptide of phage $\varrho\beta$ translated as shown in fig.19) of appreciable length (>50 aminoacids) could only be translated in the +1 frame in both phages; in the case of phage MS2 the lysis polypeptide is translated in the +1 frame as shown in fig.19 (see also section 4.2.4.2.). The -1 frame, especially in $\varrho\beta$ with its numerous termination and its few initiation codons, could hardly be expected to code for such additional minor polypeptides.

Codon usage: The genetic code provides options among the codons for all aminoacids, except for methionine and tryptophane. Choices among the synonymous codons does not affect the nature of the polypeptide produced; rather these choices may relate to the base composition of a genome, to the rate of expression of a cistron (Fiers et al., 1976; Min Jou et al., 1972) and to the secondary structure of the RNA (Min Jou et al., 1972; Hasegawa et al., 1979).

Tables 5, 6, 7 and 8 show the codons used in the maturation-, coat- and replicase cistrons of $\varrho\beta$ and MS2 and in the Al-protein cistron of $\varrho\beta$. For a few aminoacids, the choice between degenerate codons seems to be non-random. It is notable that some codons are apparently avoided: in the coat cistrons of both $\varrho\beta$ and MS2 the codon AUA for isoleucine and the codons CGG, AGA and AGG for arginine are not found. Conversely, in the other cistrons (especially in the replicase cistrons of both phages) the isoleucine codon AUA is almost as frequent as AUPy.

Table 5: Codons Found in the Maturation Cistrons of $\varrho\beta$ and MS2

	U			C			A			G			
	$\varrho\beta$		MS2	$\varrho\beta$		MS2	$\varrho\beta$		MS2	$\varrho\beta$		MS2	
U	phe	18	6	ser	7	5	tyr	8	4	cys	0	0	U
		11	10		6	6		9	12		0	3	C
	leu	10	8		1	8	oc	0	0	op	1	0	A
		3	6		3	10		0	1		9	12	G
C	leu	15	6	pro	8	5	his	9	2	arg	17	7	U
		10	9		2	5		3	3		10	6	C
		4	5		2	4	gln	4	9		2	6	A
		8	2		7	3		7	9		4	3	G
A	ile	6	1	thr	8	11	asn	14	2	ser	3	4	U
		8	8		6	5		8	15		4	3	C
		4	7		4	5	lys	8	5	arg	3	3	A
	met	4	7		2	6		8	9		2	4	G
G	val	17	8	ala	13	6	asp	19	8	gly	12	15	U
		5	7		7	12		12	5		10	6	C
		5	7		4	7	glu	9	5		2	2	A
		2	9		2	10		8	12		4	5	G

Table 6: Codons Found in the Coat Cistrons of $\varrho\beta$ and MS2

	U			C			A			G			
	$\varrho\beta$		MS2	$\varrho\beta$		MS2	$\varrho\beta$		MS2	$\varrho\beta$		MS2	
U	phe	2	1	ser	3	3	tyr	3	0	cys	1	1	U
		1	3		1	2		1	4		1	1	C
	leu	2	1		1	2	oc	0	1	op	1	0	A
		0	0		3	2		0	0		0	2	G
C	leu	2	2	pro	3	2	his	0	0	arg	4	3	U
		2	2		1	1		0	0		2	1	C
		0	2		2	2	gln	3	1		1	0	A
		6	0		2	1		5	5		0	0	G
A	ile	1	4	thr	6	4	asn	3	4	ser	2	0	U
		3	4		4	4		6	6		0	4	C
		0	0		1	0	lys	3	5	arg	0	0	A
	met	1	3		1	1		4	1		0	0	G
G	val	8	4	ala	7	5	asp	4	1	gly	3	3	U
		2	4		1	2		2	3		1	3	C
		2	3		4	6	glu	1	2		1	2	A
		1	3		3	1		4	3		2	1	G

(termination codons: amber (am), ochre (oc), opal (op))

Table 7: Codons Found in the Replicase Cistrons of $\Omega\beta$ and MS2

	U			C			A			G			
	$\Omega\beta$	MS2		$\Omega\beta$	MS2		$\Omega\beta$	MS2		$\Omega\beta$	MS2		
U	phe	13	12	ser	17	7	tyr	11	5	cys	10	5	U
		15	16		8	12		12	16		5	1	C
	leu	8	8		5	6	oc	1	0	op	0	0	A
		11	5		12	11		0	1		7	9	G
C	leu	8	7	pro	15	10	his	4	4	arg	14	11	U
		16	15		3	4		6	6		6	13	C
		8	8		2	3	gln	4	7		4	4	A
		9	7		9	9		7	8		5	8	G
A	ile	14	7	thr	7	4	asn	13	11	ser	7	4	U
		9	13		6	12		13	7		7	9	C
	met	13	12		9	8	lys	7	9	arg	7	4	A
		7	10		9	7		17	16		7	2	G
G	val	19	9	ala	15	15	asp	24	19	gly	19	19	U
		5	10		13	7		14	14		10	7	C
		4	6		10	9	glu	7	9		3	8	A
		3	7		9	11		21	13		7	10	G

Table 8: Codons Found in the A1-Protein Cistron of $\Omega\beta$

	U			C			A			G			
U	phe	6		ser	5		tyr	12		cys	3		U
		6			3			2			3		C
	leu	2			3		oc	0		op	1 (*)		A
		1			4			1			5		G
C	leu	8		pro	10		his	0		arg	8		U
		5			3			1			6		C
		1			5		gln	4			3		A
		7			9			8			1		G
A	ile	11		thr	11		asn	7		ser	4		U
		4			6			8			1		C
	met	4			4		lys	7		arg	0		A
		2			2			12			0		G
G	val	14		ala	13		asp	19		gly	12		U
		3			7			4			7		C
		2			5		glu	4			1		A
		3			5			8			4		G

(termination codons: amber (am), ochre (oc), opal (op))

(*) UGA termination codon of coat cistron (position 1744) is suppressed for readthrough-translation (see section 1.2.2.)

This fits neatly to the proposal of translational modulation given by Fiers (Fiers et al., 1976; Fiers & Grosjean, 1979): in a cistron translated with high efficiency a codon corresponding to a rare tRNA species (e.g. AUA) is avoided, while its presence in the maturation and replicase cistrons slows down the rate of translation (AUA would correspond to a more frequent tRNA species). Translational modulation therefore enhances the effect of the low rate of initiation detected for these cistrons (Gussin, 1966; Lodish & Zinder, 1966; Lodish, 1968; Robertson & Lodish, 1970; Kolakofsky & Weissmann, 1971a,b; Staples et al., 1971). It must be mentioned, however, that the codons implicated in translational modulation of MS2 (UAU for tyrosine, ACA for threonine, AGU for serine; Fiers et al., 1976) which are not used in the MS2 coat cistron, are found in the homologous gene of $\varphi\beta$. It becomes evident that more information will be required to establish the theory.

In general, degenerate codons with a purine nucleotide in the third position seem to be avoided in both phages: decreasing codon frequencies CGPy>CGPu>AGPu for arginine, GGPpy>GGPu for glycine and ACPy>ACPu for threonine are found. This becomes more obvious as one looks at the overall distribution of third position nucleotides within the codons used in the cistrons of $\varphi\beta$ and MS2 (table 9). In all $\varphi\beta$ cistrons U is the preferred third position base, while in MS2 C is found in this place. This is a very interesting feature, since changes in third positions of codons are often thought to be neutral. On one side this systematic bias reflects the base composition of the RNA: in $\varphi\beta$ RNA the most abundant nucleotide is U (29%, see table 4), the preferred third base is U (38%, see table 9). In MS2 RNA the most frequent nucleotide (26%, see table 4) and the preferred third base (31%, see table 9) is C. Yet the

Table 9: Distribution of Nucleotides at Third Positions
of the Codons Used in Q β and MS2

cistron	phage	third position nucleotide			
		U	C	A	G
maturation	Q β	41.3	26.4	15.0	17.3
	MS2	22.9	31.9	21.4	23.8
coat	Q β	39.5	20.9	15.7	23.9
	MS2	28.2	33.7	20.6	17.5
replicase	Q β	32.8	25.0	17.5	24.7
	MS2	27.2	29.8	18.5	24.5

Numbers in the table are percentages; boxes marked with heavy lines indicate the maximal percentage found for a given cistron and a given species.

Numbers have been calculated from the codons used in the cistrons of Q β and MS2 (see tables 5, 6, 7).

frequency of the most abundant third position nucleotide still exceeds its overall presence in the RNA. This coincides well with the idea of Grantham (Grantham et al., 1980a; Grantham et al., 1980b), who showed that in every species systematic differences in the choice of third position nucleotides in the codons of their mRNAs exist. The reason for this systematic and species-specific bias is not understood; however, it may relate to some new, hitherto unnoticed genetic character in a species, and it may be used as a new taxonomic criterion (Grantham et al., 1980a; Grantham et al., 1980b).

However, there still remains a puzzling question in this respect: in our case the "species" is actually *E.coli*, since both $\varrho\beta$ and MS2 phages take advantage of this same host for multiplication. The observed difference in the usage of third position nucleotides between $\varrho\beta$ and MS2 is therefore particularly noteworthy (and not readily understandable), because the difference does not reflect the properties of the translation machinery of *E.coli* (which is used by both phages).

4.2.2.2. Comparison of Coat-, Replicase- and Maturation Proteins

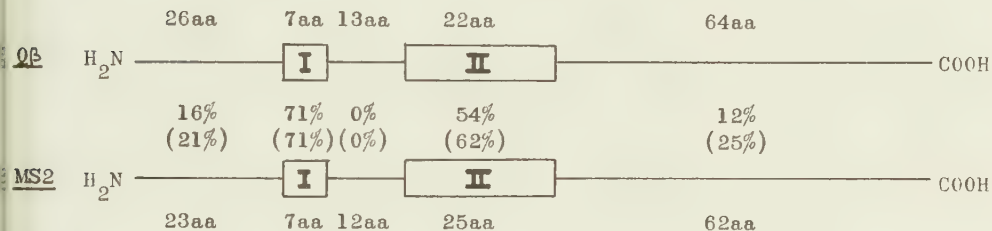
Up to now only the aminoacid sequences of MS2 and Q β coat proteins (Lin et al., 1967; Vandekerckhove et al., 1969; Maita & Konigsberg, 1971; Weber & Konigsberg, 1975; Stoll et al., 1976) and of the C-terminus of MS2 A-protein (Vandekerckhove et al., 1973) have been determined by experiment. Knowing now the nucleotide sequences of Q β and MS2 RNA and the reading frame for every gene product permits establishing the aminoacid sequences of all phage-specified proteins.

It will be shown in this section that comparison of the aminoacid sequences of homologous proteins yields further insight into the nature of the relationship between phages Q β and MS2. Comparison of aminoacid sequences is made by introducing insertions or deletions (or both) where necessary to maximize homology. Apart from homologous, i.e. identical aminoacids in the same position in the proteins compared also analogous aminoacids with similar chemical properties are considered. For aminoacid analogy the following grouping is used: aromatic (phe, trp, tyr), long chain nonpolar (leu, ile, val), short chain slightly nonpolar (ala, gly), hydroxyl group containing (thr, ser) and basic aminoacids (lys, arg) (Dayhoff et al., 1972). Furthermore, when giving numerical values for the degree of relatedness of two aminoacid sequences, aminoacid homology and analogy are indicated separately.

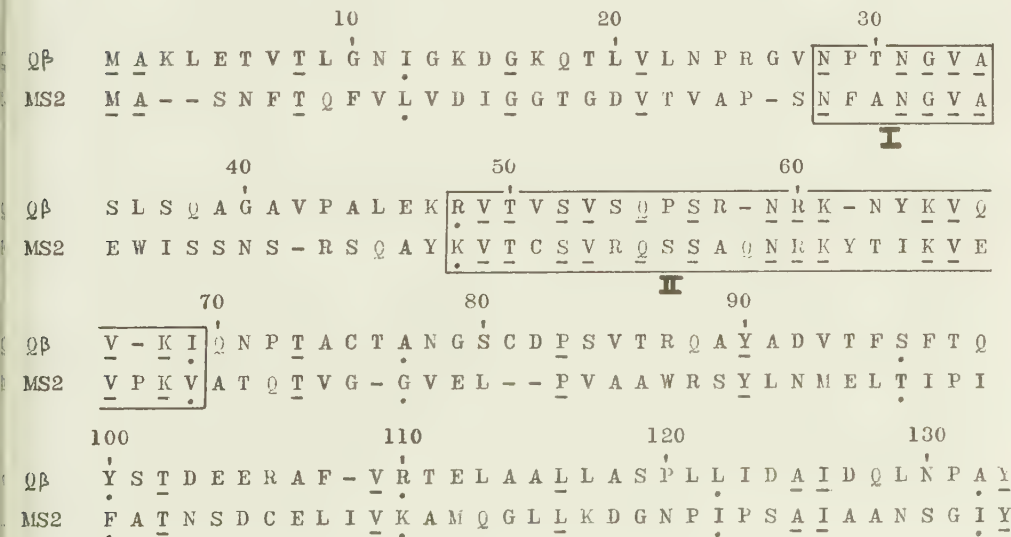
Coat proteins: The aminoacid sequences of the coat protein of phages Q β (Konigsberg et al., 1970; Maita & Konigsberg, 1971; Weber & Konigsberg, 1975; Stoll et al., 1976) and MS2 (Lin et al., 1967; Vandekerckhove et al., 1969; Weber & Konigsberg, 1975) have been determined directly and are discussed in the references given above. Therefore only the most salient features are considered in this section.

Fig.20 compares the aminoacid sequences of the coat proteins of phages $\varphi\beta$ and MS2 (the one-letter code for aminoacids is used in all the comparisons (Dayhoff et al., 1972)): both show unrelated amino- and carboxyterminal aminoacid sequences. However, two related aminoacid sequence blocks are found (blocks I and II in fig.20), comprising 25% of the complete coat protein sequence. One block (block I; 7 aminoacids) is found in the aminoterminal fourth and the other one (block II; 25 aminoacids in MS2, 22 aminoacids in $\varphi\beta$; the different length is due to three deletions in the sequence of the $\varphi\beta$ coat protein) in the central part of the coat polypeptide. Both blocks I and II show 60-70% combined aminoacid sequence homology and analogy (aminoacid deletions were taken into account as non-matching aminoacid pairs for this calculation). All other parts of the molecule have no more than 25% combined aminoacid sequence homology and analogy. It is noteworthy in this context that the larger one of the two related sequence blocks (block II) has a high content of basic residues (25%), while outside this region this value is 5-7%. This fact may relate to the dual role of coat protein to interact both with RNA and with the neighbouring coat protein molecules. But whether this cluster of conserved basic aminoacids serves as part of the interface between the coat proteins and the encapsidated RNA (which, as a general type of interaction is similar in both phages) and whether the diverged hydrophobic sequences found outside block I are responsible for protein-protein interactions (which regards individual fits between the coat protein molecules) remains to be seen. Furthermore it seems rather unlikely that the sequence of basic aminoacids is responsible for the specific interaction of the coat polypeptide with the RNA in the region preceding the replicase cistron (this interaction is required for repression of translation; see section 1.2.2.); coat proteins

(a) schematic representation of homology regions: boxes indicate segments of $\geq 50\%$ aminoacid homology; percentages refer to degree of homology for a particular segment (percentages in brackets: combined homology and analogy).



(b) aminoacid sequence: insertions and deletions are introduced to maximize homology among the sequences; homologous residues are underlined, analogous residues are marked with dots.



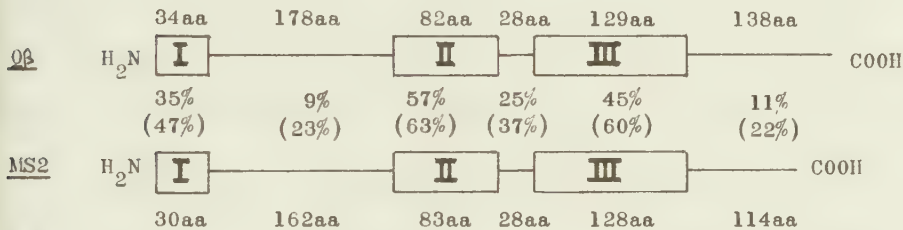
of phages R17 and MS2 from group I do not form the specific repression complexes with RNA of phage $Q\beta$ from group III, or vice versa (Bernardi & Spahr, 1972), and the RNA sequences bound by the coat proteins are quite different in group I and group III phages (Nichols, 1970; Weiner & Weber, 1971). Finally, in view of the number of insertions and deletions required to bring the identical aminoacid residues into alignment, it has not been possible to conclude that the phages in group I and III evolved from a common precursor (Weber & Konigsberg, 1975; but see following section).

Replicase proteins: Phage replicases have a rather complex structure (see section 1.2.1.); to form the active replicating enzyme in both MS2 and $Q\beta$ the products of the respective replicase cistrons (the β -subunits) combine with the same host specific proteins (ribosomal protein S1 and the elongation factors EF-Tu and EF-Ts, which are present in the enzyme complex as α -, γ - and δ -subunit). The replication mechanism, the specific recognition of the phage RNA (to exclude replication of host RNA or even of phage RNA from a different group), the function of the different replicase subunits and of the replicase associated proteins have been reviewed in more detail in sections 1.2.1. and 1.3.

In contrast to the $Q\beta$ coat protein the aminoacid sequence of the $Q\beta$ replicase β -subunit has not been determined directly (except for the first five aminoterminal residues; Weiner et al., 1972). In the case of phage MS2 the complete sequence of the replicase β -subunit was deduced already several years ago from the RNA sequence (Fiers et al., 1976). After the determination of the complete nucleotide sequence of $Q\beta$ RNA it has become possible to compare the aminoacid sequences of the β -subunits of phage MS2 and $Q\beta$ replicases, as shown in fig. 21. Comparing figs. 20, 21 and 22 it is evident that the replicase β -subunits

Figure 21: Comparison of the Aminoacid Sequences of the Replicase β -Subunit Proteins from Phages Q β and MS2

- (a) schematic representation of homology regions: boxes indicate segments of $\geq 40\%$ aminoacid homology; percentages refer to degree of homology for a particular segment (percentages in brackets: combined homology and analogy).



- (b) aminoacid sequence: insertions and deletions are introduced to maximize homology among the sequences; homologous residues are underlined, analogous residues are marked with dots.

	10	20	30	
Q β	M S K T A S S R N S L S A Q L R R A A Q T R I E V E G N L A L S I A			
MS2	M S K T T K K F N S L C I D L P R D L S L E I Y Q S - - - I A S V A			
	I			
	40	50	60	
Q β	N D L L L A Y G Q S P F N S E A E C I S F S P R F D G T P D D F R I			
MS2	- - - - - T G S G D P H S D D F T A I A Y L R D E L L T K			
	70	80	90	100
Q β	N Y L K A E I M S K Y D D F S L G I D T E A V A W E K F L A A E A E			
MS2	H P T L G S G N D E A T R R T L A I A K L R E A N G D R G Q I N R E			
	110	120	130	
Q β	C A L T N A R L T R P D Y S E D F N F S L G E S C I H M A R R K I A			
MS2	G F L D H K S L S W D P D V L Q T S I R S L I G N L L S G Y R S S L			

	140	150	160	170
Qβ	K L I G - D V P S V E G M L R H C R F S G G A T T T N N R S Y G H P			
MS2	- F G Q S T F S Q G A P M G H K L Q D A A P Y K K F A E Q A T V T P			

	180	190	200
Qβ	S F K F A L P Q A C T P R A L K Y V L A L R A S T H F D T R I S D I		
MS2	R A L R A A L L V R D Q C A A W I R H A V R Y N E S Y E F R - - L		

	210	220	230
Qβ	S P F N K A V T V P K N S K T D K C I A I E P G W N M F F Q L G I G		
MS2	V V G N G V F T V P K N N K I D R A A C K E P D M N M Y L Q K G V G		

	240	250	260	270
Qβ	G I L R D R L R C W G I D L N D Q T I N C R R A H E G S V T N N L A			
MS2	A F I R R R L K S V G I D L N D Q S I N Q R L A Q Q G S V D G S L A			

	280	II		290	300
Qβ	T V D L S A A S D - S I S L A L C E L L L P L G T F E V L M D L R S				
MS2	T I D L S S A S D D S I S D R L V W S F L P P E L Y S Y L D R I R S				

	310	320	330	340
Qβ	P K G R L P D G S V V T T E K I S S M G N G Y T F E L E S L I F A S			
MS2	H Y G - I V D G E T V R W E L F S T M G N G F T F E L E S M I F W A			

	350	360	370
Qβ	L A R S V C E I L D L D S S E V T V Y G D D I I L P S C A V P A L R		
MS2	I V K A T Q I H F G N A G T I G I - Y G D D I I C P S E I A P R V L		

	380	390	400
Qβ	E V F K Y V G F T T N T K K T F S E G P F R E S C G K H Y Y S G V D		
MS2	E A L A Y Y G F K P N L R K T F V S G L F R E S C G A H F Y R G V D		

	410	420	III		430	440
Qβ	V T P F Y I R H R I V S P A D L I L V L N N L Y R W A T I D G V W D					
MS2	V K P F Y I K K P V D N L F A L M L I L N R L R G W G V V G G M S D					

	450	460	470
Qβ	P R A H S V Y L K Y L K R P P K Q L Q R N T I P D G Y G D G A L V G		
MS2	P R L Y K V W V R L S S Q V P S M F F G G T D L A A D Y Y V V S P P		

	480	490	500	510
Qβ	S V L I N P F A K N R G W I R Y V P V I T D H T R D R E R A E L G S			
MS2	T A V S V Y T - K T P Y G R L L A D T R T S G F R L A R I A R G R K			
	520	530	540	
Qβ	Y L Y D L F S R C L S E S N D G L P L R G P S G C D S A D L F A I D			
MS2	F F S E K H D N G R Y I A W F H T G G E I T D S M K S A G Y R V I R			
	550	560	570	
Qβ	Q L I C R S N P T K I S R S S G K F D I Q Y I A C - S S - R V L A P			
MS2	T S E W L T P V P T F P - - - - - Q E C G P A S S P R			
	580	590		
Qβ	Y G V F Q G T K V A S L H E A			

of $\varphi\beta$ and MS2 show the highest degree of homology among all the various phage specific proteins. The homologous amino-acid sequences are found in three distinct blocks: one forming the aminoterminal region (block I in fig.21; residues 1-34, 34 and 30 aminoacids in $\varphi\beta$ and MS2) and the other two blocks in the central part of the protein (block II: residues 214-294, 82 and 83 aminoacids in $\varphi\beta$ and MS2; block III: residues 323-451, 129 and 128 aminoacids in $\varphi\beta$ and MS2). These blocks contain about 40% of the whole aminoacid sequence and show 35-57% homology (47-63% combined aminoacid homology and analogy). Sequences outside blocks I to III have only 9-25% aminoacid homology (22-37% combined homology and analogy). The existence of similar aminoacid sequences in $\varphi\beta$ and MS2 replicase β -subunits has been found for a small portion of the molecule (about 10%; Kappeler, 1979); the discovery of the very extensive homologies is new, however.

How do homologous and non-homologous aminoacid sequences relate to the function of the replicase β -subunits? It appears very probable that homologous regions interact with the accessory proteins recruited from the host; they might also form part of the active site for RNA polymerization since this function is called for in both $\varphi\beta$ and MS2 replicases. On the other hand, non-homologous unique aminoacid sequences might be responsible for the template specificity of the replicases, since the nucleotide sequences thought to be involved in template recognition (the replicase binding M and M-like sites; see section 4.2.3.1.) are different in $\varphi\beta$ and in MS2. The extensive aminoacid homologies also allow to solve the question whether or not the two phages are derived from a common ancestor. Some degree of homology between $\varphi\beta$ and MS2 β -subunits was to be expected in any case due to the functional requirements imposed on the aminoacid sequence of the pro-

teins (contact sites with the same E.coli proteins, RNA polymerization site). However, structural similarities due to convergent evolution from two different ancestral phage proteins would be expected to yield only very restricted regions of sequence similarity, since in principle contacts with the same proteins could be made using very different aminoacid sequences. Therefore the existence of such long regions with extensive aminoacid homology contradicts the idea of convergent development; it appears now definitely established that phages Q β and MS2 are derived from a common ancestor (see also section 4.2.4.1.).

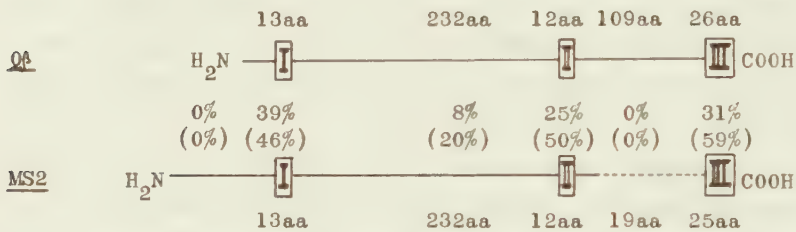
Maturation proteins: The functions of this minor capsid component (one molecule per virion) are described in section 1.1. As it proved relatively difficult to obtain sufficient amounts of pure maturation protein for sequence determination little was known about the aminoacid sequence of this polypeptide; only the aminoterminal sequences of the maturation proteins of phage R17 (8 aminoacids), of Q β (4 aminoacids) and of MS2 (43 aminoacids) have been determined directly (Weiner et al., 1972; Vandekerckhove et al., 1973).

Fig.22 shows the comparison between the aminoacid sequences of phage Q β and MS2 maturation proteins, as deduced from their nucleotide sequences. Aminoacid sequence homologies of the two proteins, in contrast to the coat and especially the replicase proteins, are scarce; they are found in three short blocks (12-26 aminoacids), occupying about 12% of the entire aminoacid sequence. These blocks show about 33% aminoacid sequence homology (52% combined homology and analogy); the residual sequences, occupying 88% of the molecule have 20% or less of combined aminoacid sequence homology and analogy.

As the homology regions found are of small size and are barely relevant statistically (Fitch, 1966a; Fitch, 1966b; McLachlan,

Figure 22: Comparison of the Aminoacid Sequences of the
Maturation Proteins from Phages Q β and MS2

- (a) schematic representation of homology regions: boxes indicate segments of $\geq 25\%$ aminoacid homology; percentages refer to degree of homology for a particular segment (percentages in brackets: combined homology and analogy).



- (b) aminoacid sequence: insertions and deletions are introduced to maximize homology among the sequences; homologous residues are underlined, analogous residues are marked with dots.

		10		20		30																														
MS2	M	R	A	F	S	T	L	D	R	E	Q	E	T	F	V	P	S	V	R	V	Y	A	D	G	E	T	E	D	N	S	F	S	L	K	Y	
		40																																		
Q β	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
MS2	R	S	N	W	T	P	G	R	F	N	S	T	G	A	K	T	K	Q	W	H	Y	P	S	P	Y	S	R	G	A	L	S	V	T	T	S	
Q β	G	L	R	F	G	A	D	N	E	I	L	N	D	F	Q	E	L	W	F	P	D	L	F	I	E	S	S	D	T	H	P	W	Y	T	L	
MS2	I	D	Q	G	A	Y	K	R	S	G	S	S	W	G	R	P	Y	E	E	K	A	G	F	G	F	S	L	D	A	A	S	C	Y	S	L	
		110																																		
Q β	K	G	R	V	L	N	A	H	L	D	D	R	L	P	N	V	G	G	R	Q	V	R	R	T	P	H	R	V	T	V	P	I	A	S	S	
MS2	F	P	V	S	Q	N	L	T	Y	I	G	V	P	Q	N	V	A	N	R	A	S	T	E	V	L	Q	K	V	T	Q	G	N	F	N	L	
Q β	G	L	R	P	V	T	T	V	Q	Y	D	P	A	A	L	S	F	L	L	N	A	R	V	D	W	D	F	G	N	G	D	S	A	N	L	
MS2	G	V	A	L	A	E	A	R	S	T	A	S	Q	L	A	T	Q	T	I	A	V	V	K	A	Y	T	A	A	R	R	G	N	W	R	Q	

	180	190	200	210
Qβ	V I N D F L F R T F A P K E F D F S N S L V P R Y T Q A F S A F N A K			
MS2	A L R Y L A L N E D R K F R S K H V A G R W I E L Q F G W L P I M S D			

	220	230	240
Qβ	Y G T M I G E G L E T I K Y L G L L L R R L R E G Y R A V K R G D L R		
MS2	I G A Y E M L T K V H L Q E F L P M R A V R Q V G T N I K L D G R L S		

	250	260	270	280
Qβ	A L R R V I Q S Y H N G K W K P A T A G N L W L E F R Y G L M P L F Y			
MS2	Y P A A N F Q T T S N I S R R I V I W F Y I N D A R L A W L S S L G I			

	290	300	310
Qβ	D I R D V M L D W Q N R H D K I Q R L L R F S V G H G E D Y V V E F D		
MS2	L N P L G I V W E K V P F S F V V D W L - - P V G N M L E G L T A P V		

	320	330	340	350
Qβ	N L Y P A V A Y F K L K G E I - T L E R R H R H G I S Y A N R E G Y A			
MS2	G C S Y M S G T V T D V I T G E S I I S V D A P Y G W T V E R Q G T A			

II

	360	370	380
Qβ	V F D N G S L R P V S D W K E L A T A F I N P H E V A W E L T P Y S F		
MS2	- K A Q I S A M H R G V Q S V W P T T G - - - - -		

	390	400	410	420
Qβ	V V D W F L N V G D I L A Q Q G Q L Y H N I D I V D G F D R R D I R L			
MS2	- - - - -			

	430	440	450
Qβ	K S F T I K G E R N G R P V N V S A S L S A V D L F Y S R L H T S N L		
MS2	- - - - -		

	460	470	480
Qβ	P F A T L D L D T T F S S F K H V L D S I F L L T Q R V K R		
MS2	- - - - A Y V K S P F - S M V H T L D A L A L I R Q R L S R		

III

1971), the maturation proteins of phages $\phi\beta$ and MS2 seem unrelated. This finding was not surprising, since only very restricted homology of the maturation proteins might be expected: the weakly homologous regions, especially at the C-terminus, are possibly involved in the interaction with the host cell (both $\phi\beta$ and MS2 phages attach to the same F-pili of E.coli; Crawford & Gesteland, 1964; Valentine & Strand, 1965). By contrast, interaction with the coat protein in the phage capsid probably requires no aminoacid sequence homology, since the coat polypeptides of phages $\phi\beta$ and MS2 differ in 75% of their aminoacid sequences. Similarly the interaction with the RNA (see section 1.1.) calls for little homology in the maturation proteins, since both RNAs of phages $\phi\beta$ and MS2 are vastly different in their nucleotide sequences. However, as long as the interactions of maturation proteins with other proteins or the RNA are not clearly defined, the explanation given above for the low degree of homology remains without any basis.

4.2.3. Comparison of Recognition Sequences

In addition to the comparison of the coding sequences ideally all the other complex functions of the RNA outlined in sections 1.2. and 1.3. should be reviewed on the basis of its nucleotide sequence. Probably the whole RNA structure is important for many functions, but for the time being the attention has to be focussed on the comparison of hitherto defined recognition sequences of the RNA. Such regions are, among others, RNA-protein contact regions (replicase- and coat protein binding sites), RNA-RNA interaction sites (ribosome binding sites) and the extracistronic nucleotide sequences shown to be required for the proper functioning of the phage (see sections 1.1 to 1.3.).

4.2.3.1. Replicase Binding Sites

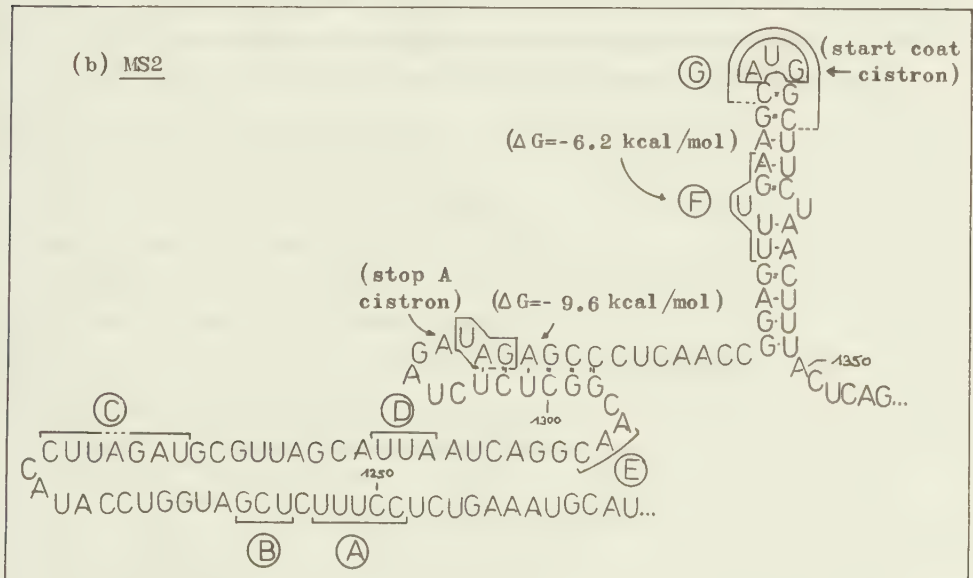
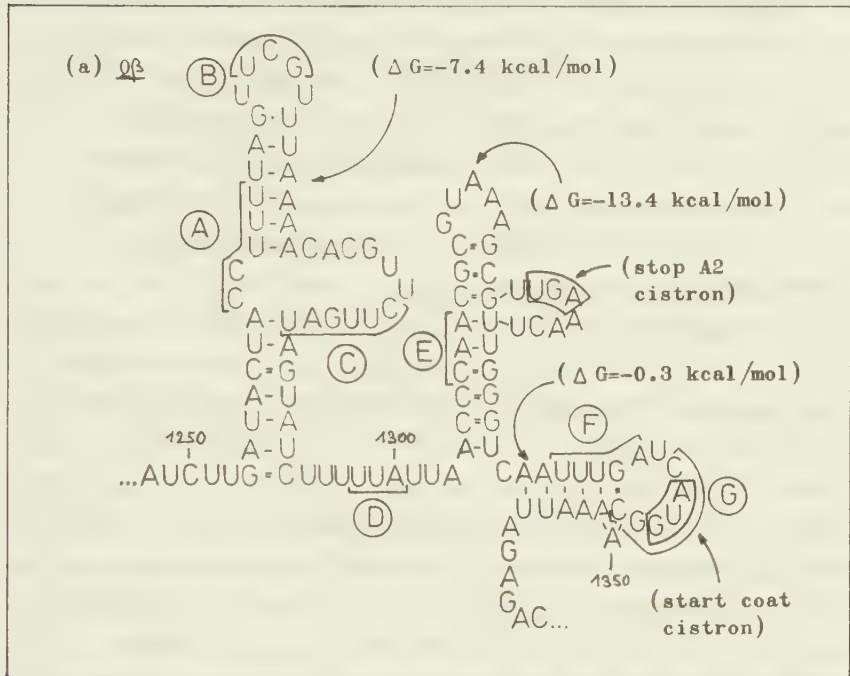
Replicase binding sites have been defined only for $Q\beta$, since MS2 replicase is exceedingly unstable and is usually only active with artificial templates (Weissmann et al., 1964; Fiers et al., 1967). Experiments with the $Q\beta$ system have established two main binding regions of $Q\beta$ replicase on the RNA, viz. the S- and the M-sites (Weber et al., 1972; Meyer, 1978). The S-site, located at the ribosome binding region of the coat cistron (nucleotides 1248-1347), is occupied by replicase even in the absence of magnesium ions. This interaction was shown to alter the functional state of $Q\beta$ RNA: from its role as a messenger it is changed into a template for replication (Kolakofsky & Weissmann, 1971; see also section 1.2.2.). In the case of MS2 the same conversion from messenger to template has to take place; it is therefore reasonable to assume that MS2 replicase binds to a functionally analogous region of the RNA to prevent further access of ribosomes to the coat cistron (Kolakofsky & Weissmann, 1971).

The second binding site is only active in the presence of magnesium ions; under such conditions $Q\beta$ replicase interacts, in addition to the S-site, with a region in the replicase cistron, defined as M-site. The M-site was found to consist of three closely spaced regions (M5: nucleotides 2546-2614, M2b: nucleotides 2638-2812, M11: nucleotides 2845-2873 (Meyer, 1978); see also fig.1). It was shown to be required for recognition of the template for replication (Meyer, 1978): RNA fragments lacking this site could not be replicated by replicase (Schwyzer, 1973). Since this site is required for specific recognition of the RNA by its replicase and the two replicase β -subunits are still rather different (despite the long homologous regions discussed in section 4.2.2.2.) it was completely open whether MS2 replicase binds to a corresponding region in MS2 RNA; if so, it

was not clear whether this region diverged from a corresponding region of a common ancestor phage together with the replicase β -subunit.

Replicase binding S-site: Comparison of the $\Omega\beta$ replicase S-site with the corresponding region in MS2 RNA (the last seventy nucleotides of the maturation cistron, the adjacent intercistronic region and the first fifteen nucleotides of the coat cistron; see fig.23) shows approximately 32% homology in the nucleotide sequences; for this calculation only completely homologous sequences larger than two nucleotides were taken into account. Homology is found in seven blocks (blocks A-G in fig. 23) of three to seven nucleotides; the homology in block C is not complete: an additional A-residue is found in the case of MS2. The significance of the homologies found is not clear, however; if the restriction of three consecutive nucleotides in register is not applied a homology of 25% is expected in any case for two entirely unrelated nucleotide sequences. For an even more specific reason the homology found might be misleading, since it may in part reflect an aminoacid homology at the C-termini of the maturation proteins of $\Omega\beta$ and MS2. As shown in section 4.2.2.2. the last 26 aminoterminal aminoacids of the maturation proteins of both phages have the highest degree of homology (31%) of the entire polypeptide (see fig.22). It is therefore unclear whether the nucleotide sequence homology in the S-site is imposed by requirements of the S-site or of the maturation protein structure. Similarly the homology regions near the initiation triplet of the coat cistron (block F in fig.23) might be required for ribosome binding (Hindley & Staples, 1969; Steitz, 1972). However, the pentanucleotide UUUGA (block F) is not part of a Shine-Dalgarno interaction sequence with 16S ribosomal RNA (such an interaction was shown to be an important feature in the specific

Figure 23: Comparison of Q β RNA Replicase Binding S-Site and a Corresponding Region in MS2 RNA



(legend fig.23)

The secondary structure for $Q\beta$ RNA replicase binding S-site is as published (Weber et al., 1972; Escarmis et al., 1978); the secondary structure for the corresponding region in MS2 RNA is as published by Fiers (Fiers et al., 1975).

Calculations of the thermodynamic stabilities of the stem-and-loop structures was made according to the published rules (Tinoco et al., 1973); the G-values given in the figure refer to the entire particular stem-and-loop structure, including bulges where present.

Sequence blocks (A) to (G) are homologous nucleotide sequences in $Q\beta$ and MS2 RNA; numbers in connection with nucleotides indicate the nucleotide positions in the genome, counting from the 5'-end.

It must be mentioned that in the case of MS2 only the secondary structure for the 3'-terminal moiety of the sequence is considered (nucleotides 1298-1394). Additional long-range interactions of nucleotides 1236-1293 with two sequences located between positions 1119-1153 and 1197-1218 and of nucleotides 1318-1322 with the sequence located between positions 950-954 have been published (Fiers et al., 1975); these long-range interactions are not shown in this figure. In the case of $Q\beta$ RNA corresponding long-range interactions are not known.

binding of ribosomes to most prokaryotic messenger RNAs (Shine & Dalgarno, 1974)). It is also noteworthy that the S-site in $Q\beta$ and the corresponding region in MS2 lie within an RNA region slightly enriched in A- and U-residues (65% A+U for $Q\beta$, 55% for MS2), indicating secondary structures melting at lower temperatures than surrounding sequences.

The secondary structures of the $Q\beta$ S-site (using the data of Weber et al. (1972) and of Escarmis et al. (1978)) and of the corresponding region in MS2 (Fiers et al., 1975) reveal three stem-loop structures for $Q\beta$ and two for MS2 as the only prominent structural features found in this region (but see legend to fig. 23). It must be mentioned that in the case of MS2 additional base pairings of nucleotides 1236-1293 with two sequences located between positions 1119-1153 and 1197-1218 and of nucleotides 1318-1322 with a sequence found between positions 950-954 have been published (Fiers et al., 1975); due to the lack of experimental data similar long-range interactions are not known for $Q\beta$ RNA (see section 4.2.1.).

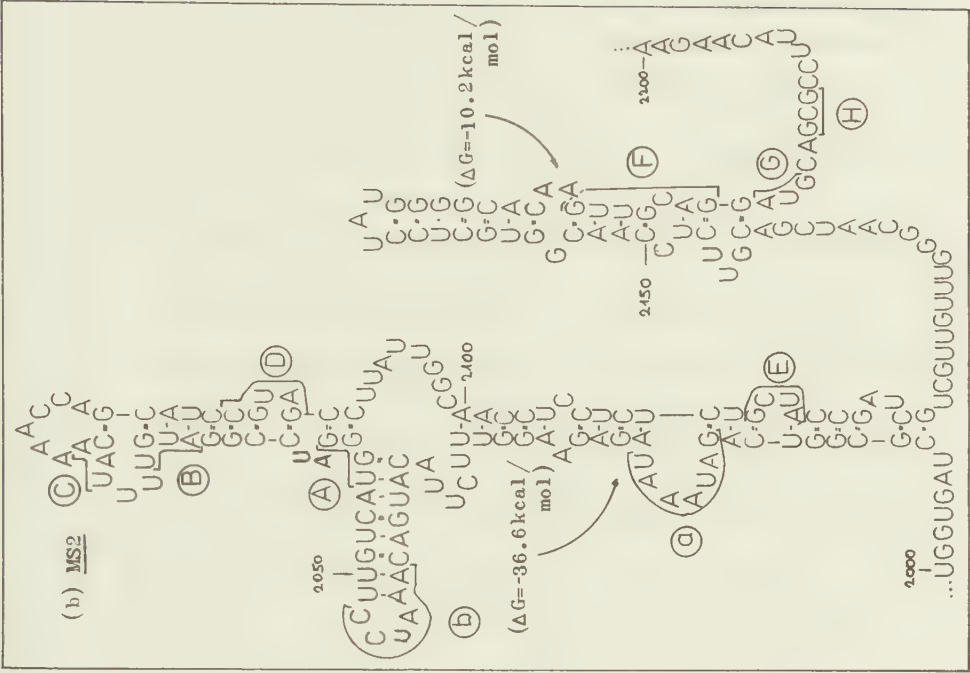
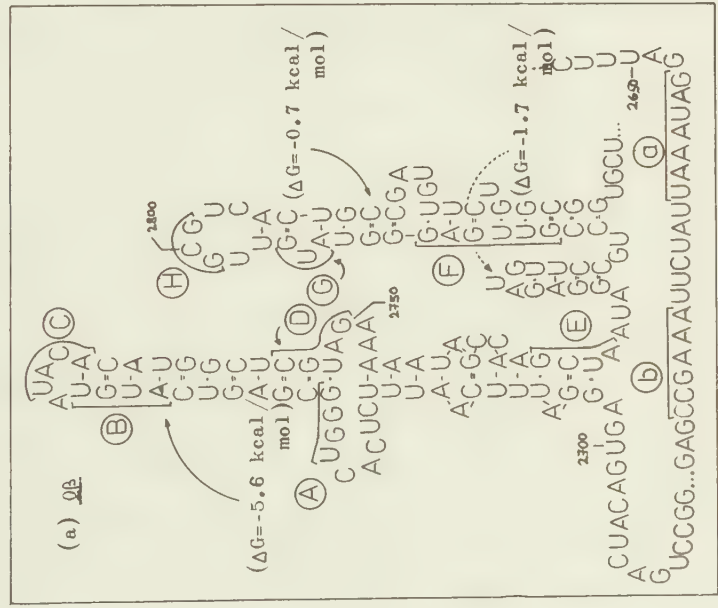
The calculated stabilities of the stem-and-loop structures seen in this region are found to be in the order of -6 to -13 kcal/mol, with the exception of the structure containing the initiation codon of the $Q\beta$ coat cistron ($G=-0.3$ kcal/mol). All stem-loop structures are therefore stable entities and must occur in more or less appreciable amounts.

It is clear at this point that, although some homology in the primary sequences of the $Q\beta$ and MS2 S-sites exists, this type of homology is misleading: the evidence for similar secondary structures in the S-sites is very weak. It remains therefore open whether replicase binds to this region due to primary or secondary structure features, or whether higher order structures are also involved in recognition and binding.

Replicase binding M-sites: For reasons outlined above the search for M-like sites in MS2 was problematic; only regions in roughly the same area of the replicase cistron (in analogy to the M-sites of $Q\beta$; see fig.1) were considered. Fig.24 shows the comparison between $Q\beta$ M-5 and M-2b sites and the corresponding region in MS2 RNA ($Q\beta$: positions 2547-2826, corresponding to nucleotides 193-473 from the start of the replicase cistron; MS2: positions 2000-2200, corresponding to nucleotides 239-439 from the start of the replicase cistron). A region similar to the M-11 site of $Q\beta$ could not be found in the RNA of MS2, at least not in this region. In terms of nucleotide sequence both $Q\beta$ M-5 and M-2b sites and the MS2 counterparts show about 20% homology by the standards explained above. The homologies are found in stretches of oligonucleotides ($Q\beta$ M-5 and the corresponding site in MS2: a hexa- and a heptanucleotide, viz. blocks a and b in fig.24; $Q\beta$ M-2b and the analogous region in MS2 RNA: eight oligonucleotides of three to six residues, viz. blocks A-H in fig.24). As to secondary structure both nucleotide sequences give rise to two prominent stem-and-loop structures; however, the structures are very different in size and in thermodynamic stability, and most of the homologous oligonucleotides are found in different locations of the proposed structures. Furthermore, since only the secondary structure in the case of MS2 is supported by experimental data structural features. In any event it is clear that very little primary and secondary structure similarity between $Q\beta$ M-5/M-11 sites and MS2 RNA exists. As mentioned above this is not surprising in view of the function of these sites.

Possibly the large differences between $Q\beta$ and MS2 M- and "M-like" sites provide an example for divergent evolution: in every protein-RNA contact area each change in protein structure was followed by a concomitant change in the nucleotide sequence, or vice versa (a "dialogue" situation between protein and RNA).

Figure 24: Comparison of 2β RNA Replicase Binding
M-Sites and a Corresponding Region in
MS2 RNA



(legend fig.24)

The secondary structure for $Q\beta$ RNA replicase binding M-site is an extended version of the structure proposed by Meyer for the M-2 site (Meyer, 1978); the structure is derived by calculation and is not supported by direct experimental evidence. The secondary structure for the corresponding region in MS2 RNA is as published (Fiers et al., 1976). Calculations of the thermodynamic stabilities of the stem-and-loop structures were made according to the published rules (Delisi & Crothers, 1971; Tinoco et al., 1973); all the ΔG -values given in the figure refer to the entire particular stem-and-loop structure, including bulges where present.

Sequence blocks (a) and (b) are homologous nucleotide sequences corresponding to the $Q\beta$ M-5 site, blocks (A) to (H) correspond to the $Q\beta$ M-2 site; numbers in connection with nucleotides indicate the nucleotide positions in the genome, counting from the 5'-end.

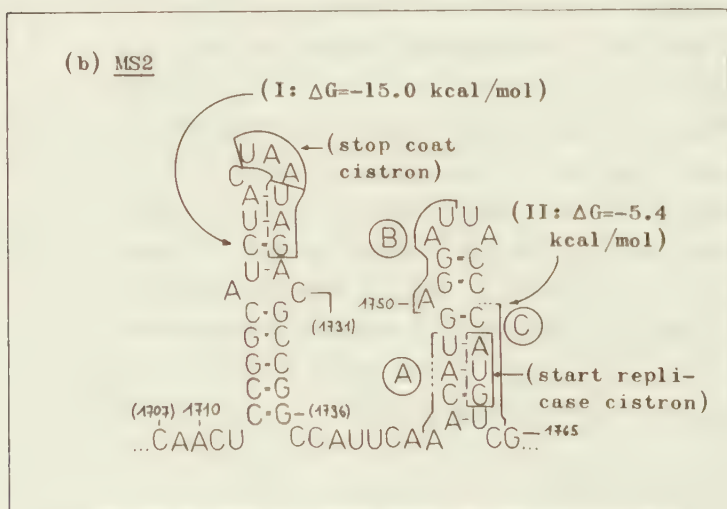
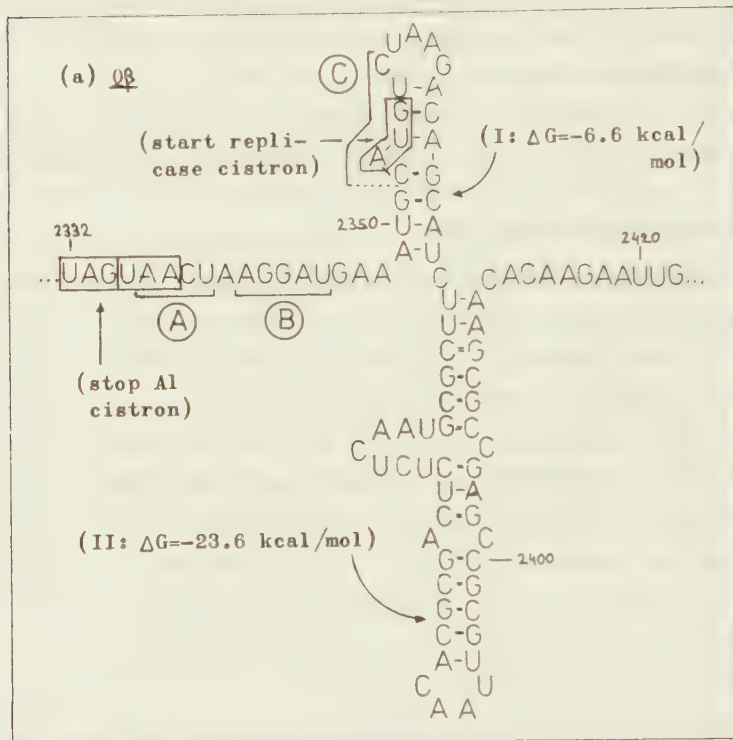
In view of the very tenuous similarity of $Q\beta$ M-sites and the corresponding RNA region of MS2 it is altogether doubtful whether MS2 replicase binds to this region. It may be possible that the specific recognition of the RNA by MS2 replicase takes place at a completely different RNA region. However, this statement does not necessarily follow from the sequence dissimilarity found.

4.2.3.2. Coat Protein Binding Sites

As mentioned in sections 1.1. and 1.2.2. the viral coat protein has, besides its function as the main component of the phage capsid, a second function as a translational repressor to stop further synthesis of replicase β -subunit molecules. The phenomenon of coat-mediated translational repression was found in MS2 and $Q\beta$ (Eggen & Nathans, 1968; Robertson et al., 1968; Weber, 1976). Although coat proteins of group I phages do not form the necessary specific complexes with the RNA of group III phages (Bernardi & Spahr, 1972), and only about 25% of the aminoacid sequences of $Q\beta$ and MS2 coat proteins are homologous (see fig.20) a comparison of the two coat protein binding sites appeared interesting; in both cases the coat protein molecules have to bind to functionally analogous regions of the RNA.

Fig.25 shows the sites, both spanning similar regions around the initiation sites of the replicase cistrons ($Q\beta$: nucleotides 2332-2422; MS2: nucleotides 1707-1765). About 23% nucleotide sequence homology between $Q\beta$ and MS2 RNAs is found in this region, using the standards explained for the case of the replicase binding S-site. Homology is found in three blocks (blocks A-C in fig.25) of four to six nucleotides; the homology in block A is rather weak, however: an additional A-residue is found in the case of MS2. As to the significance of the primary structure homology the same argument mentioned

Figure 25: Comparison of Q β RNA Coat Protein Binding Site and a Corresponding Region in MS2 RNA



(legend fig.25)

The secondary structure for the coat protein binding site of MS2 RNA is as published for phage R17 (Gralla et al., 1974); the secondary structure for the corresponding region of $\varphi\beta$ RNA is from Weber (Weber, 1976). The minimal RNA fragment bound by coat protein is loop II in MS2 (Gralla et al., 1974) and loop I in $\varphi\beta$ (Weber, 1976). In the region shown in this figure the nucleotide sequences of MS2 and R17 RNAs differ from each other by one nucleotide at position 1731 in loop I of MS2, which is probably unimportant for coat binding (position 1731: MS2 has a C-residue, while R17 RNA has a U-residue).

For positions 1707-1736 of MS2 RNA the secondary structure given is identical to the proposal of Min Jou (Min Jou et al., 1972), while for positions 1737-1765 these authors have postulated base-pairing with positions 1412-1433 of the coat cistron (Min Jou et al., 1972); the difference between the two proposed secondary structures may reflect the difference between the RNA structures of non-translated RNA and RNA with the coat cistron sequences occupied by ribosomes. In the case of $\varphi\beta$ no experimental evidence for similar long-range interactions of the coat protein binding site with sequences located in the coat cistron exists.

Calculation of the thermodynamic stabilities of the stem-and-loop structures was made according to the published rules (Tinoco et al., 1973); the ΔG -values given refer to the entire particular stem-and-loop structure, including bulges where present.

Sequence blocks (A) to (C) are homologous nucleotide sequences in $\varphi\beta$ and MS2 RNA; numbers in connection with nucleotides indicate the nucleotide position in the genome, counting from the 5'-end.

for the replicase binding S-site (see 4.2.3.1.) must be raised: the coat binding sites cover ribosome-RNA interaction sites which obviously show some sequence homologies (Shine & Dalgarno, 1974). Furthermore the homologous hexanucleotide C seen at the beginning of the replicase cistrons reflects the common sequence fmet-ser found at the N-termini of both $\phi\beta$ and MS2 replicases. Interm of secondary structure both $\phi\beta$ and MS2 coat protein binding sites form two weakly and two moderately stable stem-loop structures, as proposed by Gralla et al. (Gralla et al., 1974) for R17 and MS2 (see legend to fig.25) and by Weber for $\phi\beta$ (Weber, 1976). Again the *prima facie* nucleotide sequence homologies found prove to be very misleading: the nucleotide sequences protected by the coat proteins are of different lengths ($\phi\beta$: 91 nucleotides, MS2: 59 nucleotides), they are shifted with respect to the replicase cistron initiation region and the arrangement of the homologous nucleotide sequences within the proposed secondary structures is very different in the two RNAs.

It thus appears that the two coat binding sites are really very different. As in the replicase binding M-sites, the coat proteins and their recognition sequences on the RNA may have evolved independently in the two phages; each change in the protein structure necessitated a corresponding change in the RNA recognition sequence, or vice versa. This suggests that for the coat protein-RNA interaction the diverged parts of the proteins are involved, and it fits the observation that heterologous coat protein-RNA complex formation between group I/III phages has not been possible (Bernardi & Spahr, 1972).

4.2.3.3. Non-Translated Regions

It has been proposed that non-translated RNA sequences in phage RNAs may serve as recognition and control elements for translation and replication of the RNA (Domingo et al., 1976; Sabo et al., 1977; Taniguchi & Weissmann, 1978); they may also have an

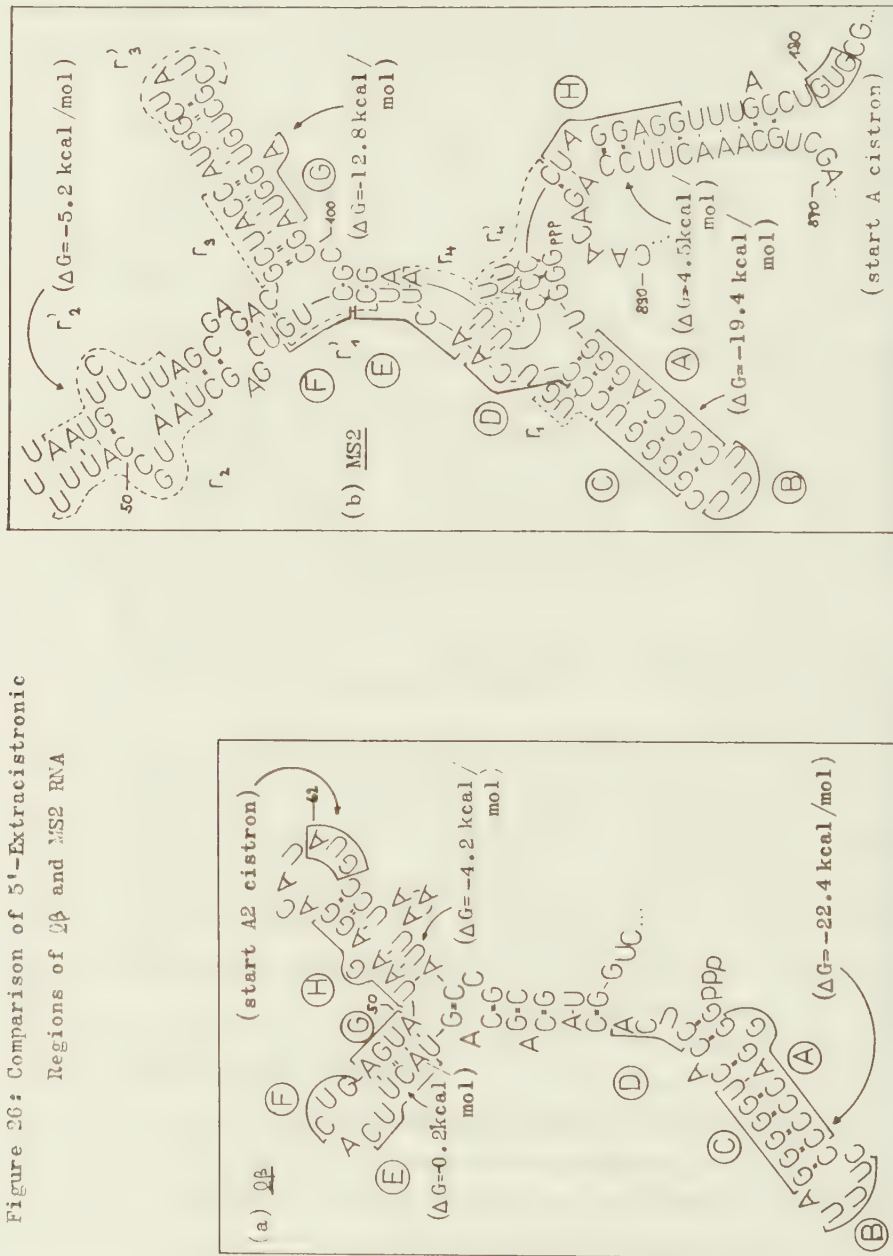
essential biological role in connection with resistance against nucleolytic degradation of the RNA, or its packaging (Weissmann et al., 1973). These functions might be carried out by the RNA folding into unique secondary structures and by providing specific binding sites for proteins or for protein-RNA complexes, such as ribosomes.

In $\varphi\beta$, point mutations in extra- or intercistronic sequences are known to affect the infectivity of the RNA, its replication rate and its interaction with ribosomes (Domingo et al., 1976; Taniguchi & Weissmann, 1978). It has furthermore been proposed that nontranslated RNA regions were less prone to point mutations during evolution than cistronic nucleotide sequences: among group I phages (MS2, R17, f2) 3-4% nucleotide variability within the cistrons was found (Nichols, 1971; Min Jou et al., 1972; Robertson & Jeppesen, 1972), while nucleotide sequences of extracistronic regions are conserved among these RNAs (Robertson & Jeppesen, 1972), and only one variant position in each of the two intercistronic regions was identified in MS2 and R17 (Cory et al., 1970; Min Jou et al., 1972).

In figs. 26 and 27 the 5'- and 3'-extracistronic RNA regions, respectively, of phages $\varphi\beta$ and MS2 are compared. Both regions contain some sequences in common. The 5'-extracistronic region of the two phages show more extensive sequence homology than the 3'-regions (5'-ends: 37 homologous nucleotides out of 61 in $\varphi\beta$ and 129 in MS2; 3'-ends: 14 homologous nucleotides out of 99 in $\varphi\beta$ and 170 in MS2).

In the 5'-extracistronic regions of phages MS2 and $\varphi\beta$ (fig. 26) one homologous sequence block of known function can be identified: the sequence UAPuGAGG preceding the start of the maturation cistron (block H), which ensures the ribosome-phage RNA contact according to the model of Shine and Dalgarno

Figure 26: Comparison of 5'-Extracistrionic Regions of β and MS2 RNA



(legend fig.26)

The secondary structure for the 5'-extracistronic region of $\Omega\beta$ RNA is derived by calculation and is not supported by direct experimental evidence; the secondary structure for the corresponding region in MS2 RNA is as published (Fiers et al., 1979).

Calculation of the thermodynamic stabilities of the stem-and-loop structures was made according to the published rules (Delisi & Crothers, 1971; Tinoco et al., 1973); all the ΔG -values given in the figures refer to the entire particular stem-and-loop structure, including bulges where present.

Sequence blocks (A) to (H) are homologous nucleotide sequences in $\Omega\beta$ and MS2 RNA; sequence blocks $r_1^{(n)}$ to $r_n^{(n)}$ are repeated oligonucleotides in MS2 RNA (the sign ' indicates the corresponding repeat oligonucleotide of identical or very similar sequence). Numbers in connection with nucleotides indicate the nucleotide positions in the genome, counting from the 5'-end.

The homologous sequence blocks (A) to (F) are according to the proposal of Iserentant & Fiers (1979); the other two homologous sequences indicated ((G) and (H)) are not given by these authors who are considering nucleotide sequences up to positions 320 for $\Omega\beta$ and 117 for MS2 RNA (Iserentant & Fiers, 1979).

(legend fig.27)

The secondary structure for the 3'-extracistronic region of $\Omega\beta$ RNA is derived by calculation and is not supported by direct experimental evidence; the secondary structure for the corresponding region in MS2 RNA is as published (Fiers et al., 1976).

Calculation of the thermodynamic stabilities of the stem-and-loop structures was made according to the published rules (Delisi & Crothers, 1971; Tinoco et al., 1973); all the ΔG -values given in the figures refer to the entire particular stem-and-loop structure, including bulges where present.

Sequence blocks (A) to (C) are homologous nucleotide sequences in $\Omega\beta$ and MS2 RNA; sequence blocks $R_1^{(n)}$ to $R_2^{(n)}$ are repeated oligonucleotides in $\Omega\beta$ and sequence blocks $r_1^{(n)}$ to $r_3^{(n)}$ are repeated oligonucleotides in MS2 RNA (the sign ' indicates the corresponding repeat oligonucleotide of identical or very similar sequence, the sign " denotes a third copy of the repeat nucleotide). Numbers in connection with nucleotides indicate the nucleotide positions in the genome, counting from the 5'-end.

(Shine & Dalgarno, 1974). For phage MS2 an actual complex of this region with the 3'-end of 16S ribosomal RNA has been demonstrated (Steitz & Jakes, 1975). The functions of all the other homologous sequence blocks in the extracistronic regions remain unknown.

In both 5'- and 3'-terminal regions somewhat similar secondary structures for $\varphi\beta$ and MS2 RNA are proposed: in general the same number of stem-and-loop structures is found within each respective region for both phages; additionally in MS2 the stems are longer than in $\varphi\beta$ and two extra hairpin structures are found in the 3'-terminal region of MS2 RNA. Some homologous oligonucleotide blocks are located within a similar secondary structure environment (5'-end: oligonucleotides A, C, G and H form the stem and B the loop of stem-and-loop structures in both phages; 3'-end: oligonucleotide C forms the stem of a stem-and-loop structure found in both phages). The distribution of the other homologous sequences with respect to the secondary structure seems to be different.

One of the most intriguing features found in the extracistronic sequences are the apparent repeats which consist of identical or very similar nucleotide sequences: they seem to be responsible for the additional stem-and-loop structures or their extension, as especially seen in MS2. Although some of these apparent repeats (5'-end: oligonucleotides r_1-r_4' for MS2; 3'-end: oligonucleotides R_1-R_2'' for $\varphi\beta$, r_1-r_5' for MS2) might be present by chance, their existence suggests a possible tendency in the evolution of the extracistronic regions of MS2: the overall secondary structure seems to be retained while hairpins may be added and the stems of stem-and-loop structures may be enlarged by sequence duplications; functionally important sequences are kept in relatively constant secondary structure environments. A different hypothesis in

this context is that nucleotides 44-71 in the 5'-extracistronic region of MS2 RNA show good sequence homology (71%) with nucleotides 100-127 in the $\Omega\beta$ maturation cistron; this may suggest that the extracistronic region of MS2 was enlarged and the initiation site of the maturation protein shifted during evolution (Weissmann et al., 1973). Finally, it is obvious that the occurrence of such sequence repeats is more frequent in the extracistronic than in the coding RNA regions; it has to be noted, however, that extensive nucleotide repeats of another type are found in the A1-cistron of phage $\Omega\beta$ (see section 4.2.4.1.).

The intercistronic regions were compared already above in their functions as replicase and coat binding sites (see sections 4.2.3.1. and 4.2.3.2.). As ribosome binding sites they have been evaluated before (Steitz, 1975; Lewin, 1977) and need not be discussed in this context.

4.2.4. Minor Gene Products of Bacteriophages $\Omega\beta$ and MS2

Apart from the three "standard" polypeptides (maturation-, coat- and replicase β -subunit proteins) also minor gene products are produced in "unconventional" ways under the direction of the phage RNAs (see section 1.2.2.). In phage $\Omega\beta$ the A1 protein is synthesized by reading through the UGA termination codon of the coat cistron into the adjacent sequence of approximately 600 nucleotides (see section 1.2.2.). Translation is at 2% of the efficiency of the coat cistron translation (Weiner & Weber, 1973). In phage MS2 the lysis protein is translated out-of-phase relative to both the coat and the replicase cistrons (Atkins et al., 1979; Model et al., 1979). Here translation starts out of the normal reading frame near the end of the coat cistron across the coat-replicase intercistronic region into the beginning of the replicase cistron (see section 4. .4.2.). Translation occurs at less

than 5% of the efficiency of the coat cistron translation (Atkins et al., 1979).

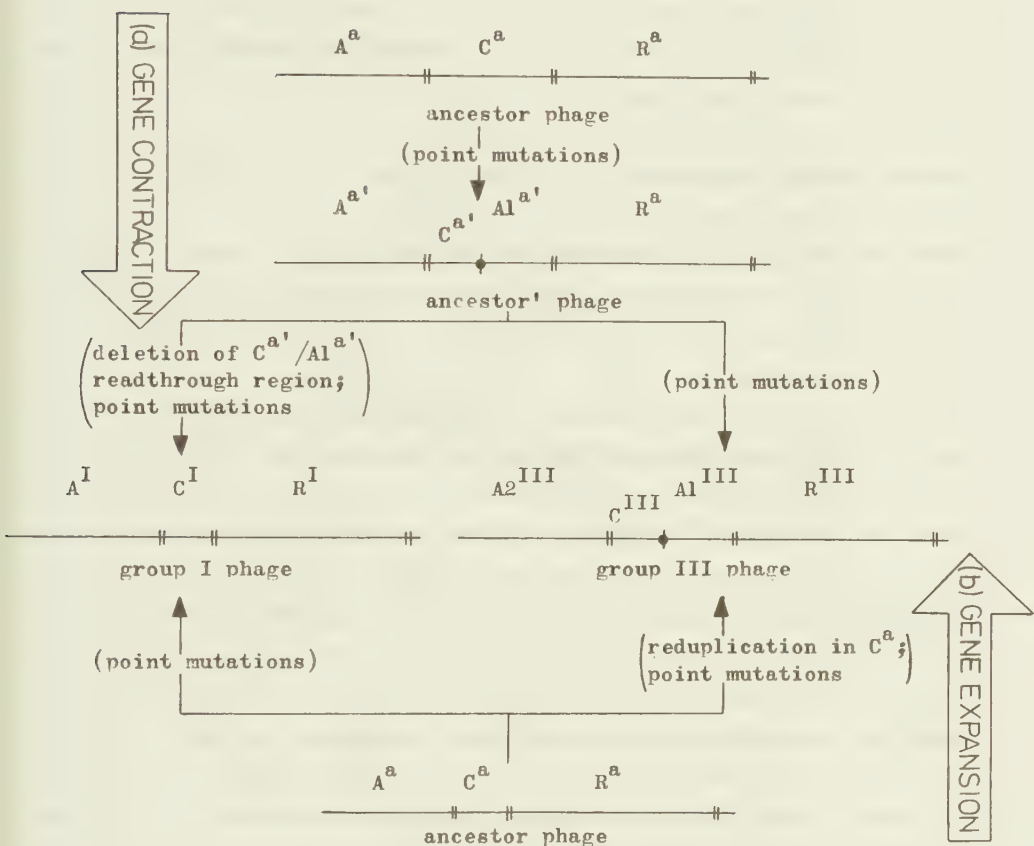
Although both minor gene products of $\varphi\beta$ and MS2 are essential for the viability of these phages their functions are poorly understood. However, by analysis of the RNA coding for these minor gene products some insight into relationships and into possible evolutionary pathways of the phages $\varphi\beta$ and MS2 can be gained as outlined in the following sections.

4.2.4.1. Evolution of $\varphi\beta$ Al Protein

In the course of this work the relatedness of phages $\varphi\beta$ and MS2 which has been suggested previously (Konigsberg et al., 1970; Robertson & Jeppesen, 1972; Hofstetter et al., 1974), has become firmly established as discussed in section 4.2.2.2. One of the most striking differences between the genomes of phages $\varphi\beta$ and MS2 is their different size, which is mainly due to the coat-Al readthrough region (see section 1.1.) of $\varphi\beta$ not present in MS2. To explain this difference Hofstetter et al. (Hofstetter et al., 1974) have proposed possible schemes for the evolution of the combined coat/Al cistron in $\varphi\beta$ from a gene found in a hypothetical common ancestor of group I and III phages. With the complete nucleotide sequences of $\varphi\beta$ and MS2 RNA at hand it has now become possible to test these proposals.

Fig.28 (scheme (a)) shows evolution by gene contraction, which involves deletion of a long nucleotide sequence. According to this proposal the ancestral phage contained a large coat cistron. In a first step a mutational event within this sequence led to a conditional termination codon: only a few copies of the long coat protein were made following this event, and the novel short protein produced in normal high amounts has taken over most of the functions of the original coat polypeptide,

Figure 28: Evolution of Coat-Al Cistrons from Hypothetical Common Ancestor Phage Cistrons



The scheme is redrawn from Hofstetter et al. (Hofstetter et al., 1974).

Cistron symbols: maturation (A, A2), Al protein (Al), replicase (R); the superscripts a, a', I, III refer to ancestor, ancestor', group I or group III phage cistrons, respectively.

Termination symbols: unconditional ($\#$), conditional ($\#$).

which was necessary only in a few copies per virion. This stage would correspond to the present stage of phages like λ . Finally, the need for the long coat protein was obviated altogether, corresponding to the situation of present group I phages. According to this scheme the evolution was dictated by economy, first reducing the amount of protein to be synthesized per phage particle, and second reducing the phage genome length.

Mechanistically such a deletion within the ancestral coat cistron could have come about by replicase jumping from one strand at the base of a stem-loop RNA structure to the other strand, without copying the looped-out sequence. Of course no noticeable trace of such an event could be found in the RNA of recent phages.

Fig.28 (scheme (b)) shows evolution by gene expansion. According to this scheme the replicase of the ancestor phage at one time copied twice a nucleotide sequence within the ancestral coat cistron (C^A). Consequently the longer cistron gave rise to a longer protein with improved properties, which initially replaced the old short coat protein. In the further course of evolution (or even during the sequence reduplication event) a conditional termination codon was created by mutation within this cistron. This would correspond to the first step in the gene contraction pathway discussed above and would result in economizing protein synthesis, with all the implications mentioned.

Mechanistically the sequence reduplication could arise when replicase, after copying a stem-loop RNA sequence, jumps back to the initial RNA strand of the stem and copies the loop once again. In principle, in the course of time several such events could occur giving rise to multiple reduplications of the sequence within a stem-loop structure. The main feature

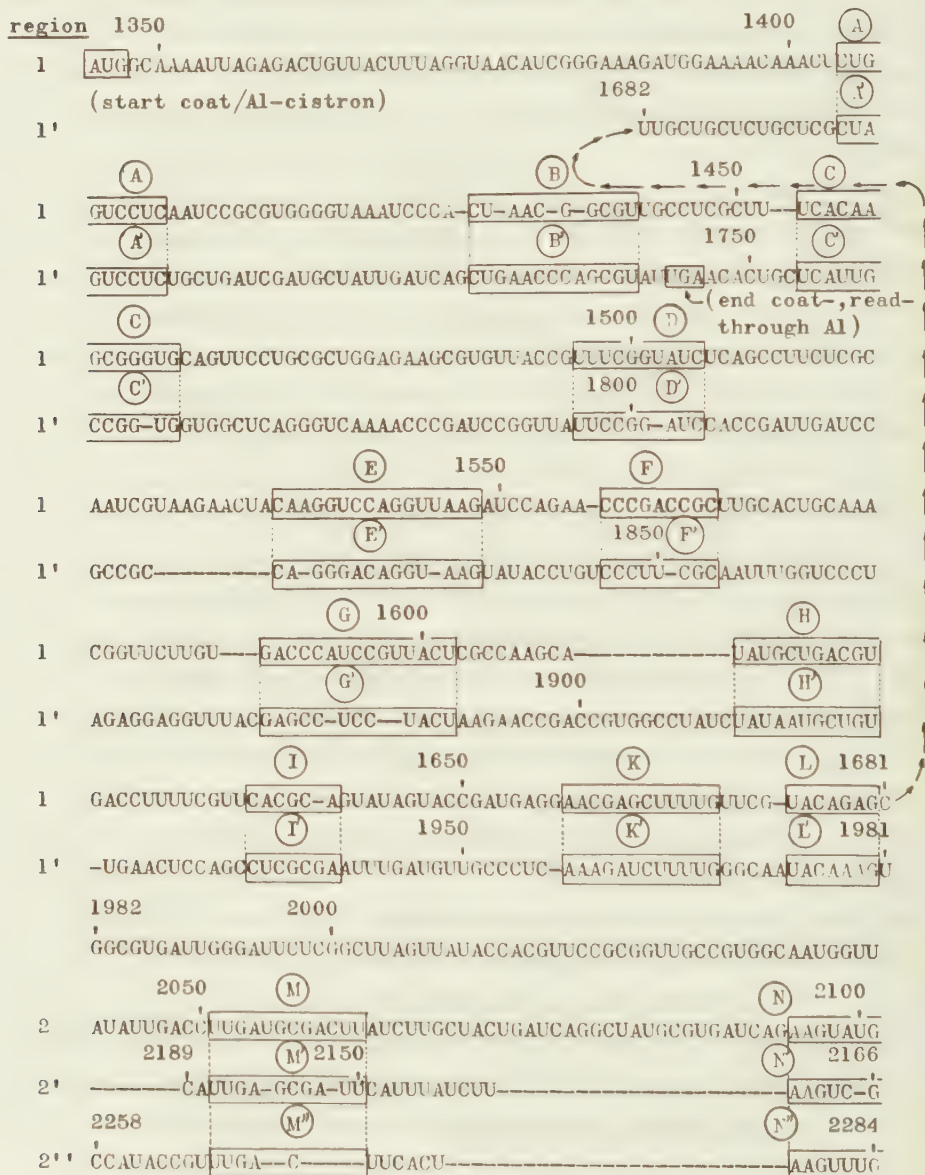
which this scheme predicts is that prominent repeats (or at least quasi-repeats of similar nucleotide sequence) be found in the coat-A1 readthrough region of phage $\varphi\beta$.

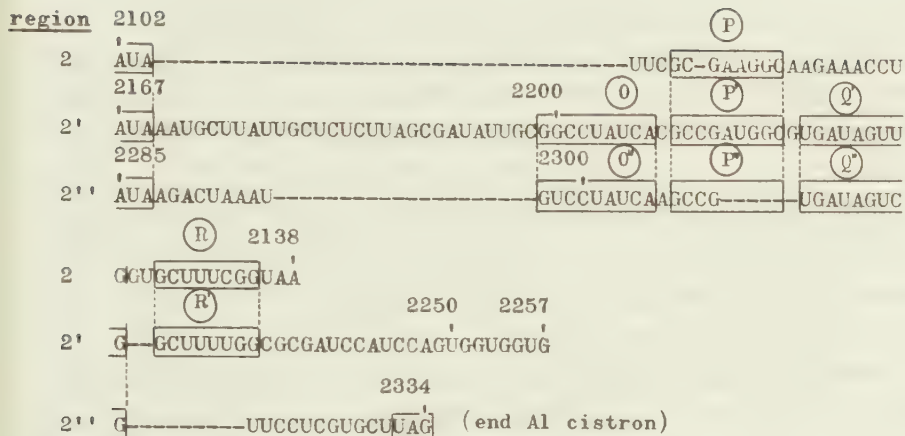
How do the expansion- and contraction pathways compare with previous experimental observations? In vitro, only deletions and no duplications of phage RNA sequences have been documented. Phage $\varphi\beta$ RNA, after extensive in vitro replication, was found to lack a large segment of its genome (nearly one fourth to one third of the entire genome, the deletion reaching from the end of the A2 cistron to the beginning of the replicase cistron (Sabo et al., 1977)). Other authors have found similar deletions in $\varphi\beta$ RNA after prolonged in vitro replication: up to 85% of the $\varphi\beta$ genome was lost in such experiments (Mills et al., 1967). As a general rule viruses seem to economize the amount of genetic material as much as possible: several regions of the single stranded DNA phages ϕ X174 (Barrell et al., 1976) and G4 (Godson et al., 1978) are used twice in different reading frames, and in the small animal DNA virus SV40 the same genome regions are used to produce different messengers and hence different proteins by various splicing pathways (Fiers et al., 1978; Reddy et al., 1978). Furthermore, group I RNA phages with their smaller genome and shorter replication times are more viable than $\varphi\beta$ and can therefore be assumed to represent a higher stage of phage evolution. A priori evolution by gene contraction seemed therefore more likely than by gene expansion.

What does the nucleotide sequence of phage $\varphi\beta$ RNA reveal? In fig. 29 the sequence of the $\varphi\beta$ coat-A1 cistron is represented in the form of blocks aligned to show repeat regions: region 1 from nucleotide positions 1404-1680, region 1' from positions 1697-1980 (each region about 280-290 nucleotides in length), followed by two shorter regions (regions 2 and 2'; positions

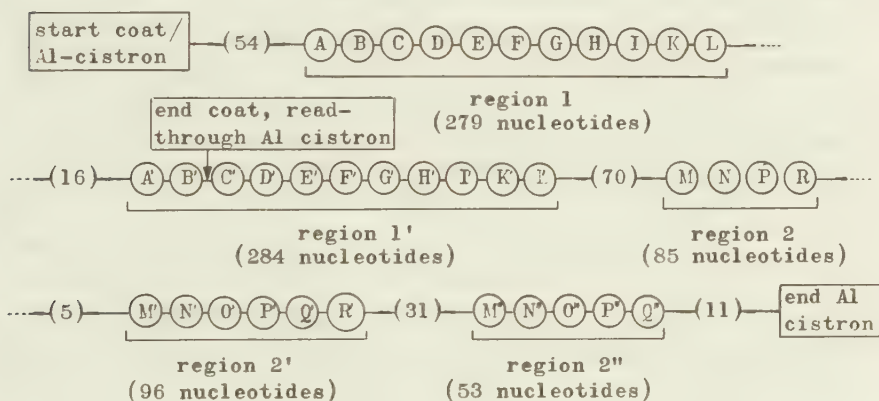
Figure 29: Repeated Nucleotide Sequences in the Coat-Al Cistron of Bacteriophage 2β

(a) sequence comparison within 2β coat/Al-cistron:





(b) schematic repeat structure:



(A) to (Q) are the oligonucleotides forming the repeat regions 1 to 2''; for their actual nucleotide sequence see scheme (a).

Position numbers in the Q β genome are given above the nucleotide sequence; numbers in brackets in scheme (b) indicate the number of nucleotides between repeat sequence regions 1 to 2''.

2051-2135 and 2141-2235, respectively; about 90 nucleotides in length). A third, shorter region (2'', positions 2267-2320, about 50 nucleotides in length) is found further downstream. These regions cover about 80% of the complete coat-Al cistron. Regions 1 and 1' are quite similar in their nucleotide sequences; regions 2, 2' and 2'' do not seem to be related to regions 1 or 1'. However, they seem to be related among themselves: region 2'' appears to share some sequences with regions 2 and 2', but 2'' is rather distantly related to regions 2 and 2'. Although these repeated regions are far from being exact copies of each other they nevertheless seem to be closely related: homologous sequences (3-15 nucleotides, showing on average 80% sequence homology) are interspersed with oligonucleotide tracts (3-27 residues) having much lower sequence homology. Nevertheless, the fact that over a length of about 280 nucleotides (regions 1 and 1') these highly homologous segments are found arranged in the same order and with about the same spacings between them strongly supports the idea of statistically relevant repeats. For the shorter repeats further downstream (regions 2, 2' and perhaps 2'') the evidence is questionable.

The existence of such "quasi-repeats" supports the idea of the $\varrho\beta$ coat-Al cistron having evolved from an ancestral coat cistron by sequence duplication, i.e. by gene expansion. The coat cistron of a common ancestor phage may have had a nucleotide sequence intermediate between region 1 of $\varrho\beta$ and the initial 200-250 nucleotides of the MS2 coat cistron. In a relatively recent duplication event regions 1 and 1' may have arisen. Since regions 1 and 1' do not seem to be related to regions 2, 2' and 2'', but only among themselves, the regions of the 1 and 2 "series" have most probably arisen by duplications of different sequences. The better homology between regions 1 and 1' (as compared to 2 and 2') suggests that the multiple

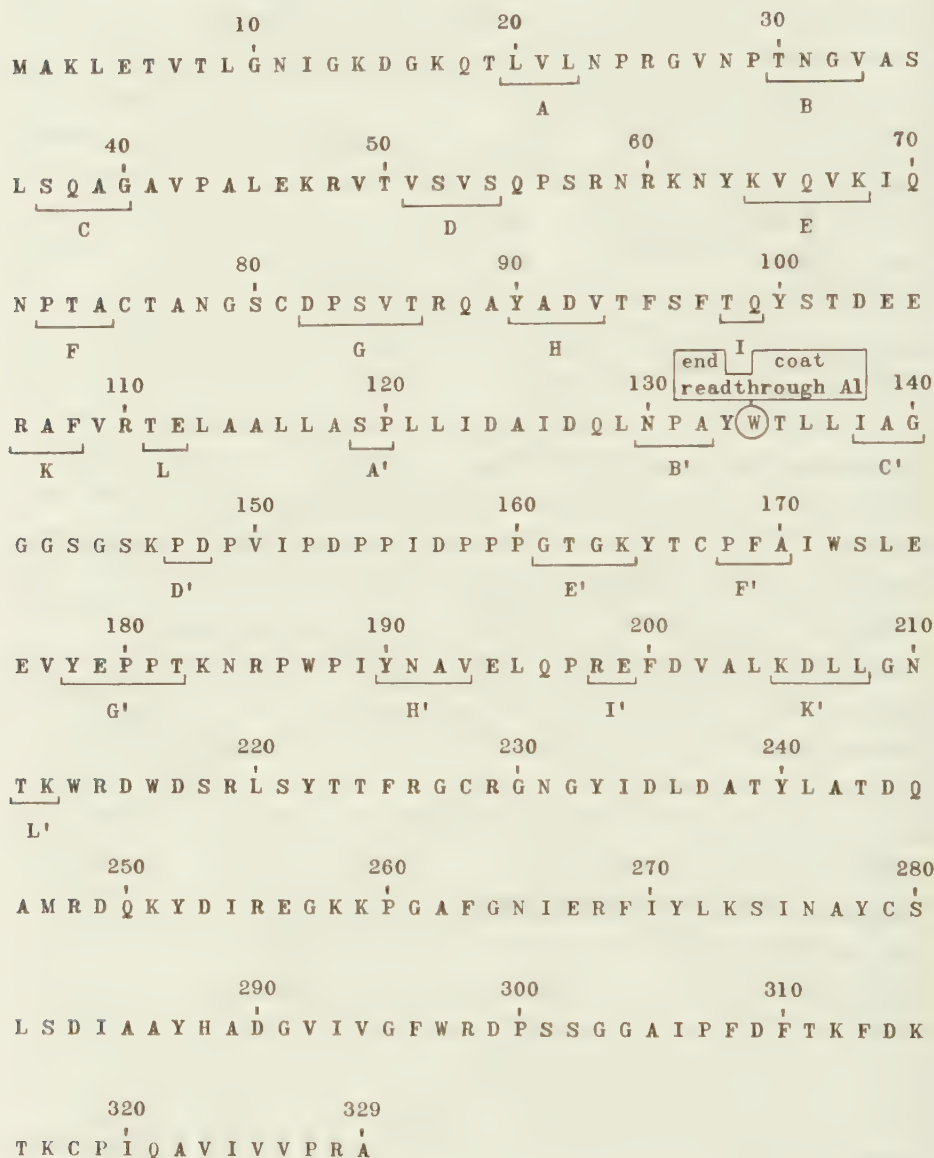
duplication events leading to regions 2, 2' and perhaps 2'' have taken place before the duplication step leading to regions 1 and 1'; regions 2, 2' and 2'' have diverged further from each other by acquiring more point mutations than regions 1 and 1', which had less time to do so.

It is not clear where the original coat termination codon of the ancestor phage was located; it would have been lost in the course of evolution and has been replaced by a new UGA termination signal in $Q\beta$ (through a point mutation or even during the duplication event leading to regions 1 and 1'). This termination codon is now found between oligonucleotides B' and C' of region 1' (see fig.29).

Although the model presented here is still rather tentative, it explains the sequence data found. The rather clear evidence for the duplication of an ancestral RNA sequence to yield regions 1 and 1' in $Q\beta$ is nevertheless surprising: even in small bacteriophages, which were thought to restrict their genome size as much as possible, genome amplification by duplicating nucleotide sequences seems to take place, a fact long known for both prokaryotic and eukaryotic organisms (Brown et al., 1972; Botchan, 1974; Ghosal & Saedler, 1978). However, small duplications in phage RNA seem to occur quite frequently, as suggested by the repeats noted in the extracistronic regions of phages $Q\beta$ and especially MS2 (see section 4.2.3.3.).

Fig.30 shows that the sequence duplications in the A1 cistron are not reflected in the aminoacid sequence of the $Q\beta$ A1 protein. This contrasts with the finding that in eukaryotic genomes repeats in the nucleotide sequence may be reflected exactly in the aminoacid sequence of the corresponding protein (e.g. the constant regions of IgG heavy chains). In RNA phages, economizing their genetic material, this type of repeat is probably useless: only those nucleotide sequence "duplications" giving rise to a different protein will become genetically fixed.

Figure 30: Aminoacid Sequence of the A1 Protein of Bacteriophage Q β



The aminoacid sequence is deduced from the nucleotide sequence; oligopeptides A-L and A'-L' correspond to the repeated nucleotide sequences in the coat-A1 cistron in region I and I' (see fig.29).

4.2.4.2. Comparison of a Hypothetical Lysis Protein in Phage ϕ X174 with the Actual Lysis Protein in Phage MS2

Several proteins encoded out of phase relative to other genes have been identified previously in the small DNA phages ϕ X174 (Barrell et al., 1976) and G4 (Godson et al., 1978). The same situation is found for the RNA phages MS2 and f2 (Atkins et al., 1979; Model et al., 1979), where a short protein of 75 aminoacids, required for lysis of the host cell, has been found; it arises by translation in the +1 frame relative to both the coat and the replicase cistrons. Translation begins at the end of the coat cistron, crosses the coat-replicase intercistronic region and stops after reading for a short distance into the replicase cistron. This protein escaped previous detection in vivo since it is only a minor translation product. However, in mammalian cell-free extracts programmed with MS2 or f2 RNA this protein could finally be found and identified; the same protein is also made in an E.coli cell-free system, but only in much lower amounts (Atkins et al., 1979). Its identification as an essential function involved in host cell lysis rests on the characterization of a UGA nonsense mutant of phage f2 (mutant op3; Horiuchi, 1975; Model et al., 1979).

Two features of the lysis proteins of MS2 and of the totally unrelated DNA phages ϕ X174 and G4 are noteworthy: they are small polypeptides (MS2/lysis: 75 aminoacids, ϕ X174/E: 95 aminoacids, G4/E: 94 aminoacids), and they have a strongly hydrophobic carboxyterminal sequence. Within this hydrophobic region the oligopeptide leu-leu-leu-ser-leu is found in the lysis proteins of MS2, ϕ X174 and G4; this stretch of aminoacids corresponds to the nucleotide sequence most highly conserved in ϕ X174 and G4 DNA (Godson et al., 1978), but its occurrence in the MS2 lysis protein may be fortuitous.

It is interesting to speculate about the origins of the lysis proteins. Possibly they have arisen by accidental translation out of the main reading frame. The protein thus produced must have conferred some selective advantage to the organism, and later on this out-of-frame translation became genetically fixed. From the evolutionary point of view cistrons translated out of phase relative to cistrons specifying major gene products are probably a late addition to the previously existing genetic makeup.

Although no essential lysis function has been demonstrated in phage ϕ up to this moment, it was nevertheless tempting to look for a polypeptide with properties similar to the MS2 lysis protein. Since in ϕ such a protein has not been demonstrated experimentally the search for it was carried out exclusively on the basis of the known ϕ RNA sequence.

To show the possibilities for a polypeptide in ϕ coded out of frame relative to the major gene products we recall fig.19 (section 4.2.2.1.). "Open" nucleotide stretches following an AUG initiation codon and devoid of termination codons are only seen in the +1 frame, while in the -1 frame, where practically no initiation codons are found, no such possibility is seen. Inspection of the initiation/termination map in fig.19 shows that the sites corresponding to translation of the lysis protein in MS2 are blocked in ϕ : due to the occurrence of multiple termination codons the sequences from the end of the coat cistron into the adjacent Al-readthrough region or from the end of the Al cistron into the replicase cistron cannot be translated in the +1 frame.

However, translation in this frame is possible in the Al-readthrough sequence from nucleotides 1946 (AUG codon) to 2168 (UAA codon). This is practically the only possibility to form a polypeptide of reasonable length and of the expected hy-

drophobic and basic properties. The nucleotide sequence of this region would code for a polypeptide of 74 aminoacids; fig.31 shows the comparison between the MS2 lysis protein and this hypothetical out-of-frame translation product of phage Q β . The two polypeptides have two similar aminoacid blocks (for the definition of aminoacid sequence similarity see section 4.2.2.2.): block I consists of 26 aminoacids (35% of the total lysis protein sequence); in the proposed Q β protein it is located in the central part, whereas in the MS2 lysis polypeptide this block is to be found in the carboxyterminal third (27% aminoacid sequence homology, 61% combined homology and analogy). The other, shorter sequence (block II) consists of 11 aminoacids in Q β and of 9 aminoacids in MS2; it is located in both cases at the actual carboxyterminal end of the polypeptides (30% aminoacid sequence homology, 50% combined homology and analogy). Outside blocks I and II less than 25% combined sequence homology and analogy is found. In contrast to the MS2 and ϕ X174 lysis proteins, which lack glycine, the putative Q β polypeptide has three glycine residues; in the MS2 lysis protein the absence of glycine residues was thought to be related to the folding of the polypeptide into the specific α -helix required for its integration into the host cell membrane (Atkins et al., 1979; Beremand & Blumenthal, 1979). On the other hand, both MS2 and Q β proteins have a strikingly hydrophobic region in the carboxyterminal half of the molecule; instead of the sequence leu-leu-leu-ser-leu which forms part of this region and which is found in ϕ X174, G4 and MS2 lysis proteins, the suggested Q β protein has in its place an oligopeptide leu-ile-leu-leu-leu in a similar location. Contrary to the MS2 lysis protein, which possesses a distinctly basic N-terminal sequence, the Q β polypeptide would be rather hydrophobic in this region. It is not clear whether the

charge distribution in the putative $Q\beta$ polypeptide would allow at all its integration into the host cell membrane, a property which is presumably indispensable for the functioning of this protein (Beremand & Blumenthal, 1979). Another argument against the existence of this $Q\beta$ lysis protein is the total absence of a Shine-Dalgarno type ribosome interaction region (see section 4.2.3.1. and 4.2.3.2.; Shine & Dalgarno, 1974) preceding the initiation triplet for this putative lysis protein cistron. Further experiments would be needed to show actual synthesis of such a protein in $Q\beta$ and to prove its function in host cell lysis; yet the arguments discussed above render the actual existence of the putative $Q\beta$ lysis protein rather improbable.

It may be noted in this context, that no obvious aminoacid homologies between the MS2 lysis protein and the coat or A1 proteins of $Q\beta$ were found to exist (see figs. 20, 30 and 31). This may argue against lytic activity of $Q\beta$ coat protein (Zinder & Lyons, 1968) or, more importantly, for $Q\beta$ A1 protein, which is absent in the MS2 system and which could hence possibly fulfil over the lytic function in phage $Q\beta$. However, a lytic function may be exerted in this case by a completely different protein structure or even by a totally different general mechanism.

4.2.4.3. Hybrid Proteins Arising by Ribosomal Frameshifting

Atkins and coworkers have found that in a cell-free translation system from E.coli programmed with either MS2 or $Q\beta$ RNA various proteins are synthesized which can only be explained in terms of frame-shifting events (Atkins et al., 1979). High levels of such frameshift products are favoured if natural E. coli tRNA species (most notably $\text{ser}^{\text{AGU}}_{\text{AGC}}$ tRNA (=ser3tRNA) or $\text{thr}^{\text{ACU}}_{\text{ACC}}$ tRNA (=thr major tRNA)) are added in high concentrations relative to all other tRNAs. Specific competitor tRNAs interfering with such frameshifts have been characterized (see below).

These frameshift proteins appear even to be synthesized under "physiological" tRNA conditions, albeit in much lower amounts. In the case of MS2 two major classes of frameshift proteins are found: polypeptides related to the viral replicase protein (molecular weight 66'000-67'000, as opposed to the 62'000 of the wildtype replicase) and to the viral coat protein (molecular weight 16'000-20'000; compare with the 14'000 of the wildtype coat protein). These related proteins arise by ribosomes initiating in the standard cistrons and later shifting into a different reading frame. Translation can be carried on across termination signals in the original reading frame into adjacent nucleotide sequences, and hybrid proteins are made. One of the most interesting cases in MS2 is the coat-related polypeptide "5" (Atkins et al., 1979; Beremand & Blumenthal, 1979). It is probably 59 aminoacids longer than the coat protein and it has been identified as a coat-lysis hybrid; according to the available evidence it is synthesized by ribosomes correctly initiating at the start of the coat gene and shifting near its end into the +1 frame, the actual reading frame of the MS2 lysis protein. Synthesis of this hybrid product is enhanced by the addition of high concentrations of glNtRNA and by pro minor tRNA and is specifically competed by ala-1B tRNA. Interestingly a protein with similar characteristics has been found in vivo in a membrane fraction of infected cells (Beremand & Blumenthal, 1979); therefore it seems to be synthesized under natural conditions as well. Since the lysis mutant op3 can be complemented by coat mutants (Model et al., 1979), polypeptide 5 is not strictly essential, but it might facilitate host cell lysis.

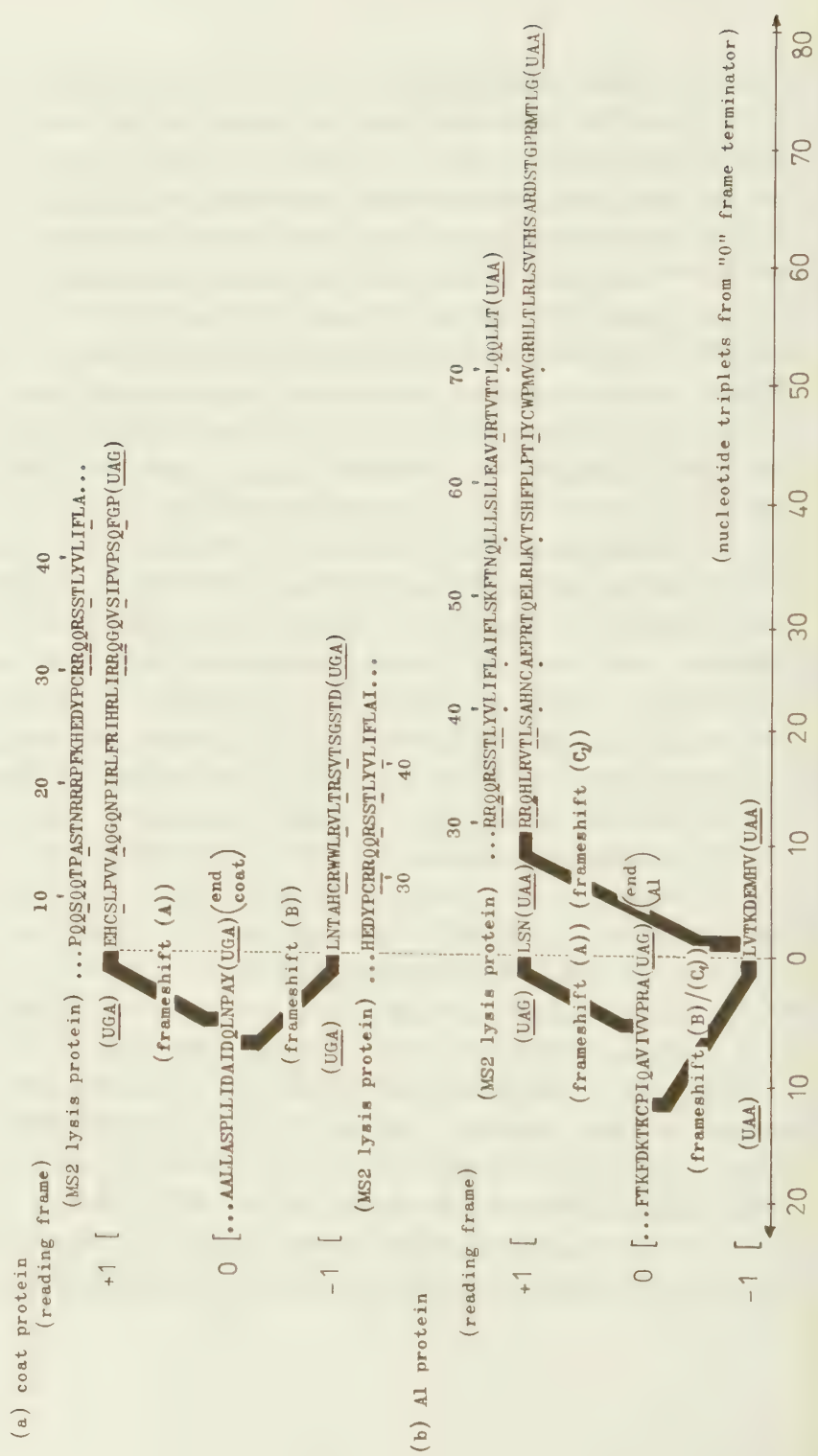
Are similar proteins found in Q β ? With Q β RNA as template in the cell-free protein synthesizing system addition of ser-3 tRNA or thr-major tRNA causes the enhanced production of specific non-standard polypeptides (Atkins et al., 1979). Synthesis of

these proteins is competed by the addition of ala-1B tRNA or of pro-minor tRNA. Hence the competitor effects are not only found with MS2 RNA as template. However, with $Q\beta$ RNA as template and under conditions promoting frameshifts in all MS2 cistrons no synthesis of a protein larger than $Q\beta$ replicase was observed. From the present evidence it is unclear from which proteins the $Q\beta$ frameshift products are derived, however.

Fig.32(a) shows possible frameshift proteins derived from the $Q\beta$ coat cistron. Invoking a +1 frameshift between nucleotides 1733 and 1743 (frameshift (A)) a protein longer than the wild-type coat polypeptide (+40 aminoacids) is to be expected. This situation is analogous to the one of the coat-lysis hybrid protein in MS2 (Atkins et al., 1979; see also above), but the aminoacid homology between the additional 40 aminoacids at the carboxyterminus of this hypothetical protein and the MS2 lysis polypeptide is as low as 17%. Such a protein is not very likely to have lytic function. Making use of a -1 frameshift between nucleotides 1725 and 1743 (frameshift (B)) a protein again longer than the coat polypeptide is expected (+23 aminoacids). This case is analogous to the coat-derived hybrid polypeptides "6" and "7" of MS2 (Atkins et al., 1979); the low aminoacid homology (only 23%) of the carboxyterminal 23 aminoacids with the MS2 lysis protein is probably fortuitous, and this product is not likely to have a lytic function, either.

Fig.32(b) shows possible frameshift proteins derived from the $Q\beta$ Al cistron. Protein (A) (arising by frameshift (A) into the +1 frame between nucleotides 2318 and 2331) would be only three aminoacids longer than Al. Protein (B), only nine aminoacids longer than Al, could be made by shifting into the -1 frame (frameshift (B), between nucleotides 2296 and 2331). It seems excluded that such a short elongation could lead to a lysis function. Only by a double frameshift (frameshifts (C_1) and (C_2)):

Figure 32: Frameshift Model for Possible Coat- and A1-Related Hybrid Polypeptides of Bacteriophage Q β



(legend fig.32)

The dashed vertical line represents the "0" frame translation terminator. The polypeptides are represented using the one-letter code for aminoacids (Dayhoff, 1972), with the imprecisely delimited switch points drawn as solid bars; the different frameshifts (denoted by capital letters interrupting the frameshift bars) and the resulting various hybrid polypeptides are explained in the text. The relevant termination codons are drawn in brackets (underlined) in the respective reading frames. For sequence comparison the relevant portions of the MS2 lysis protein are aligned with the various Q β hybrid polypeptides; alignment was made to maximize aminoacid sequence homologies. Homologous aminoacids are underlined, analogous residues are marked with dots below them. Numbers in connection with the MS2 lysis protein refer to the position of the aminoacids, counting from the N-terminal formyl-methionine.

two consecutive -1 frameshifts between nucleotides 2296 and 2331, as for protein (B), and between nucleotides 2349 and 2357, leading into the +1 frame of the replicase cistron) the protein (C) could be synthesized. This protein would be 75 aminoacids longer than A1 and would show 12% aminoacid homology with the last 25 carboxyterminal residues of MS2 lysis polypeptide; however, this homology could only reflect the homology of the N-terminal parts of $\Omega\beta$ and MS2 replicase proteins (see fig.21). Apart from the low homology, which makes a lytic function for this protein rather unlikely, it seems excluded that a double frameshifting event could be as efficient as to give rise to a reasonable amount of a protein, particularly since readthrough into the A1 region is anyway rather unfrequent (see section 1.2.2.). In fact this situation is only discussed in such detail to show that in $\Omega\beta$, unlike in MS2, a protein arising from frameshifting and reading into the replicase cistron can practically be excluded.

As a final possibility a hybrid polypeptide consisting of part of the $\Omega\beta$ A1 protein fused to the hypothetical lysis protein discussed in section 4.2.4.2. will be considered; this possibility involves a shift into the +1 frame after position 1866 during the translation of the A1-readthrough region. This hybrid polypeptide would show the reasonable sequence homology with the MS2 lysis protein for the 77 C-terminal aminoacids, as discussed in section 4.2.4.2. and as shown in fig.31. In contrast to the model given in section 4.2.4.2. the hybrid polypeptide discussed here would obviate the requirement for a ribosome recognition sequence around position 2050 (which was not found), since this model takes advantage of the high level of initiation found for the coat cistron. But whether this hybrid protein would show lytic activity and whether it would be made in sufficient amount (readthrough into the A1-region and translational frameshift would be required) remains a matter of conjecture.

REFERENCES

- Atkins J.F., Steitz J.A., Anderson C.W. & Model P. (1979). *Cell* 18:247.
- August J.T., Banerjee A.K., Eoyang L., Franze de Fernandez M.T., Hori K., Kuo C.H., Rensing U. & Shapiro L. (1968). Cold Spring Harbor Symp. Quant. Biol. 33:73.
- Bandle E.F. (1973). Ph.D. thesis, University of Zürich.
- Banerjee A.K., Rensing U. & August J.T. (1969). *J. Mol. Biol.* 71:657.
- Barrell B.G., Air G.M. & Hutchison C.A. (1976). *Natur* 264:34.
- Batschelet E., Domingo E. & Weissmann C. (1976). *Gene* 1:27.
- Beremand M.N. & Blumenthal T. (1979). *Cell* 18:257.
- Bernardi A. & Spahr P. (1972). *Proc. Nat. Acad. Sci. USA* 69:3033.
- Billeter M.A., Dahlberg J.E., Goodman H.M., Hindley J. & Weissmann C. (1969a). Cold Spring Harbor Symp. Quant. Biol. 34:635.
- Billeter M.A., Dahlberg J.E., Goodman H.M., Hindley J. & Weissmann C. (1969b). *Nature* 224:1083.
- Billeter M.A. (1978). *J. Biol. Chem.* 253:8381.
- Boedtker H. (1968). *J. Mol. Biol.* 35:61.
- Boedtker H. (1971). *Biochim. Biophys. Acta* 240:448.
- Botchan M.R. (1974). *Nature* 251:228.
- Brown D.D., Wensink P.C. & Jordan E. (1972). *J. Mol. Biol.* 94:51.
- Brownlee G.G., Sanger F. & Barrell B.G. (1967). *Nature* 215:735.
- Brownlee G.G., Sanger F. & Barrell B.G. (1968). *J. Mol. Biol.* 34:379.
- Brownlee G.G. (1972). In "Determination of Sequences in RNA", North-Holland, Amsterdam.
- Carmichael G.G., Weber K., Niveleau A. & Wahba A. (1975). *J. Biol. Chem.* 250:3607.
- Carmichael G.G., Landers T.A. & Weber K. (1976). *J. Biol. Chem.* 251:2744.
- Chang A.C.Y. & Cohen S.N. (1978). *J. Bacteriol.* 134:1141.
- Coffin J.M., Parsons J.T., Rymo L., Haroz R.K. & Weissmann C. (1974). *J. Mol Biol.* 86: 373.
- Contreras R., Vandenberghe A., Volckaert G., Min Jou W. & Fiers W. (1972). *FEBS Lett.* 24:339.

- Cory S., Spahr P.F. & Adams J.M. (1970). Cold Spring Harbor Symp. Quant. Biol. 35:1.
- Cory S., Adams J.M., Spahr P.F. & Rensing U. (1972). J. Mol. Biol. 63:41.
- Crawford E.M. & Gesteland R.F. (1964). Virology 22:165.
- Currier T.C. & Nester E.W. (1976). Anal. Bioch. 76:431.
- Dayhoff M.O. (1972). In "Atlas of Protein Sequences and Structure", vol. 5, (ed. M.O.Dayhoff), National Biomedical Research Foundation, Washington D.C.
- Delisi C. & Crothers D.M. (1971). Proc. Nat. Acad. Sci. USA 68:2682.
- De Wachter R., Merregaert J., Vandenberhe A., Contreras R. & Fiers W. (1971). Eur. J. Biochem. 22:400.
- De Wachter R. & Fiers W. (1972). Anal. Biochem. 49:184.
- Domingo E., Flavell R.A. & Weissmann C. (1976). Gene 1:3.
- Domingo E., Sabo D., Taniguchi T. & Weissmann C. (1978). Cell 13:735.
- Edgell M. & Ginoza W. (1965). Virology 27:23.
- Eggen K. & Nathans D. (1968). Fed. Proc. 27:645.
- Eggen K. & Nathans D. (1969). J. Mol. Biol. 39:293.
- Eikhom T.S. & Spiegelman S. (1967). Proc. Nat. Acad. Sci. USA 57:1833.
- Escarmis C., Sastry P.A. & Billeter M.A. (1978). J. Biol. Chem. 253:8390.
- Fedoroff N.V. & Zinder N.D. (1971). Proc. Nat. Acad. Sci. USA 68:1838.
- Feix G., Pollet R. & Weissmann C. (1968). Proc. Nat. Acad. Sci. USA 59:145.
- Feix G. & Hake H. (1975). Biochim. Biophys. Res. Commun. 65:503.
- Fiers W. (1975). In "RNA Phages" (ed N.D.Zinder), pp.353-396, Cold Spring Harbor Laboratory.
- Fiers W. (1979). In "Comprehensive Virology" (ed. H.Fraenkel-Conrat & R.R.Wagner), vol. 13, p.69. Plenum Press, New York.
- Fiers W., Verplancke H. & Van Styvendaele B. (1967). In " Regulation of Nucleic Acid and Protein Biosynthesis" (eds. V.V.Konigsberger & L.Bosch), vol. 10, p.154. Elsevier, New York.

- Fiers W., Contreras R., Duerinck F., Haegeman G., Merregaert J., Min Jou W., Raeymakers A., Volckaert G., Ysebaert M., Van de Kerckhove J., Nolf F. & Van Montagu M. (1975). *Nature* 256:273.
- Fiers W., Contreras R., Duerinck F., Haegeman G., Iserentant D., Merregaert J., Min Jou W., Molemans F., Raeymakers A., Van den Berghe A., Volckaert G. & Ysebaert M. (1976). *Nature* 260:500.
- Fiers W., Contreras R., Haegeman G., Rogiers R., Van de Voorde A., Van Heuverswyn H., Van Herreweghe J., Volckaert G. & Ysebaert M. (1978). *Nature* 273:113.
- Fiers W. & Grosjean H. (1979). *Nature* 277:328.
- Fischbach F.A., Harrison P.M. & Anderegg J.W. (1965). *J. Mol. Biol.* 13:638.
- Fitch W.M. (1966a). *J. Mol. Biol.* 16:9.
- Fitch W.M. (1966b). *J. Mol. Biol.* 16:17.
- Flavell R.A., Sabo D.L., Bandle E.F. & Weissmann C. (1974). *J. Mol. Biol.* 89:255.
- Franze de Fernandez M.T., Hayward W.S. & August J.T. (1972). *J. Biol. Chem.* 247:842.
- Garwes D., Sillero A. & Ochoa S. (1969). *Biochim. Biophys. Acta* 186:166.
- Gilvarg C., Bollum F.J. & Weissmann C. (1975). *Proc. Nat. Acad. Sci. USA* 72:428.
- Ghosal D. & Saedler H. (1978). *Nature* 275:611.
- Godson G.N. & Sinsheimer R.L. (1967). *J. Mol. Biol.* 23:495.
- Godson G.N., Barrell B.G., Staden R. & Fiddes J.C. (1978). *Nature* 276:236.
- Goodman H.M., Billeter M.A., Hindley J. & Weissmann C. (1970). *Proc. Nat. Acad. Sci. USA* 67:921.
- Gould H.J. (1967). *J. Mol. Biol.* 29:307.
- Gralla J. & Crothers D.M. (1973). *J. Mol. Biol.* 73:497.
- Gralla J., Steitz J.A. & Crothers D.M. (1974). *Nature* 248:204.
- Grantham R., Gautier C., Gouy M., Mercier R. & Pavé A. (1980a). *Nucleic Acid Res.* 8:49.

- Grantham R., Gautier C. & Gouy M. (1980b). *Nucleic Acid Res.* 8:1893.
- Groner Y., Scheps R., Kamen R., Kolakofsky D. & Revel M. (1972). *Nature New Biol.* 239:19.
- Grosjean H., Sankoff D., Min Jou W., Fiers W. & Cedergren R.J. (1978). *J. Mol. Evol.* 12:113.
- Grunstein M. & Hogness D. (1975). *Proc. Nat. Acad. Sci. USA* 72:3961.
- Gussin G.N. (1966). *J. Mol. Biol.* 21: 435.
- Haruna I. & Spiegelman S. (1965a). *Proc. Nat. Acad. Sci. USA* 54:579.
- Haruna I. & Spiegelman S. (1965b). *Proc. Nat. Acad. Sci. USA* 54:1189.
- Hasegawa M., Yasunaga T. & Miyata T. (1979). *Nucleic Acids Res.* 7:2073.
- Hindley J. & Staples D.H. (1969). *Nature* 224:964.
- Hindley J., Staples D.H., Billeter M.A. & Weissmann C. (1970). *Proc. Nat Acad. Sci. USA* 67:1180.
- Hirt B. (1967). *J. Mol. Biol.* 26:365.
- Hofstetter H., Monstein H.J. & Weissmann C. (1974). *Biochim. Biophys. Acta* 374:238.
- Hofstetter H., Schamböck A., van den Berg J. & Weissmann C. (1976). *Biochim. Biophys. Acta* 454:587.
- Hori K., Eoyang L., Banerjee A.K. & August J.T. (1967). *Proc. Nat. Acad. Sci. USA* 57:1790.
- Hori K., Harada K. & Kuwano M. (1974). *J. Mol. Biol.* 86:699.
- Horiuchi K. (1975). In "RNA Phages" (ed. N.D.Zinder), pp.29-50, Cold Spring Harbor Laboratory.
- Horiuchi K., Lodish H.F. & Zinder N.D. (1966). *Virology* 28:438.
- Horiuchi K. & Matsubashi S. (1970). *Virology* 42:49.
- Horiuchi K., Webster R.E. & Matsubashi S. (1971). *Virology* 45:429.
- Iserentant D. & Fiers W. (1979). *Eur. J. Biochem.* 102:595.
- Jeppesen P.G.N., Steitz J.A., Gesteland R.F. & Spahr F. (1970). *Nature* 226:230.
- Kacian D.L., Mills D.R., Kramer F.R. & Spiegelman S. (1972). *Proc. Nat. Acad. Sci. USA* 69:3038.

- Kamen R. (1970). *Nature* 228:527.
- Kamen R. (1975). In "RNA Phages" (ed. N.D.Zinder), pp.203-234, Cold Spring Harbor Laboratory.
- Kamen R., Kondo M., Römer W. & Weissmann C. (1972). *Eur. J. Biochem.* 31:44.
- Kappeler M. (1979). Diploma thesis, University of Zürich.
- Kolakofsky D. & Weissmann C. (1971a). *Nature New Biol.* 231:42.
- Kolakofsky D. & Weissmann C. (1971b). *Biochim. Biophys. Acta* 246:596.
- Kondo M., Gallerani R. & Weissmann C. (1970). *Nature* 228:525.
- Konigsberg W., Maita T., Katze J. & Weber K. (1970). *Nature* 227:271.
- Kozak M. & Nathans D. (1971). *Nature New Biol.* 234:209.
- Kozak M. & Nathans D. (1972). *Bacteriol. Rev.* 36:109.
- Krahn P.M. & Paranchych W. (1971). *Virology* 43:533.
- Krahn P.M., O'Callaghan R.J. & Paranchych W. (1972). *Virology* 47:628.
- Küppers B. & Sumper M. (1975). *Proc. Nat. Acad. Sci. USA* 72:2640.
- Landers T.A., Blumenthal T. & Weber K. (1974). *J. Biol. Chem.* 249:5801.
- Lewin B. (1977). In "Gene Expression" vol. 3, pp.802-811, Wiley, New York.
- Lin J.Y., Tsung C.M. & Fraenkel-Conrat H. (1967). *J. Mol. Biol.* 24:1.
- Loeb T. & Zinder N.D. (1961). *Proc. Nat. Acad. Sci. USA* 47:282.
- Loening U.E. (1967). *Biochem. J.* 102:251.
- Lodish H.F. (1968). *Nature* 220:345.
- Lodish H.F. (1971). *J. Mol. Biol.* 56:627.
- Lodish H.F., Horiuchi K. & Zinder N.D. (1965). *Virology* 27:139.
- Lodish H.F. & Zinder N.D. (1966). *J. Mol. Biol.* 19:333.
- Lodish H.F. & Robertson H.D. (1969). *Cold Spring Harbor Symp. Quant. Biol.* 34:655.
- Maita T. & Konigsberg W. (1971). *J. Biol. Chem.* 246:5003.
- Mandel M. & Higa A. (1970). *J. Mol. Biol.* 53:453.
- Maniatis T., Sim K.G., Efstradiatis A. & Kafatos F.C. (1976). *Cell* 8:163.
- Mantei N., Schwarzstein M., Streuli M., Panem S., Nagata S. & Weissmann C. (1980). *Gene* 10:1.

- Maxam A.M. & Gilbert W. (1977). *Proc. Nat. Acad. Sci. USA* 74:560.
- Maxam A.M. & Gilbert W. (1980). *Meth. Enzymol.* 65:499.
- McLachlan A.D. (1971). *J. Mol. Biol.* 61:409.
- Mekler P., van den Berg J. & Billeter M.A. (1977). *Experientia* 33:824.
- Mekler P., Marti D., Kappeler M., Jaussi R., Kadulova E., Levanon A., Schmid A. & Billeter M. (1978). *Experientia* 34:947.
- Meyer F. (1978). Ph.D. thesis, University of Zürich.
- Mills D.R., Peterson R.L. & Spiegelman S. (1967). *Proc. Nat. Acad. Sci. USA* 58:217.
- Mills D.R., Bishop D.H.L. & Spiegelman S. (1968). *Proc. Nat. Acad. Sci. USA* 60:713.
- Mills D.R., Kramer F.R. & Spiegelman S. (1973). *Science* 180:916.
- Mills D.R., Kramer F.R., Dobkin C., Nishihara T. & Spiegelman S. (1975). *Proc. Nat. Acad. Sci. USA* 72:4252.
- Min Jou W., Hindley J. & Fiers W. (1968). *Arch. Int. Physiol. Biochim.* 76:194.
- Min Jou W., Haegeman G. & Fiers W. (1971). *FEBS Lett.* 13:105.
- Min Jou W., Haegeman G., Ysebaert M. & Fiers W. (1972). *Nature* 237:82.
- Miyake T., Haruna I., Shibah T., Itoh Y.H., Yamane K. & Watanabe I. (1971). *Proc. Nat. Acad. Sci. USA* 68:2022.
- Model P., Webster R.E. & Zinder N.D. (1979). *Cell* 18:235.
- Moore C.H., Farron F., Bohnert D. & Weissmann C. (1971). *Nature New Biol.* 234:204.
- Müller W., Weber H., Meyer F. & Weissmann C. (1978). *J. Mol. Biol.* 124:343.
- Nichols J.L. (1970). *Nature* 225:147.
- Nichols J.L. (1971). *Biochim. Biophys. Acta* 228:678.
- Okada Y., Ohno T. & Haruna I. (1971). *Virology* 43:69.
- Paranchych W., Ainsworth S.K., Dick A.J. & Krahn P.M. (1971). *Virology* 45:615.
- Peacock A.C. & Dingman C.W. (1967). *Bioch.* 6:1818.

- Prives C.L. & Silverman P.M. (1972). *J. Mol. Biol.* 71:657.
- Reddy V.B., Thimmappaya B., Dhar R., Subramanian K.N., Zain B.S., Pan J., Ghosh P.K., Celma M.L. & Weissmann S.M. (1978). *Science* 200:494.
- Rensing U. & August J.T. (1969). *Nature* 224:853.
- Roberts R.J. (1979). *Meth. Enzymol.* 68:27.
- Roberts J.W. & Steitz J.A. (1967). *Proc. Nat. Acad. Sci. USA* 58:1416.
- Robertson H.D., Webster R.E. & Zinder N.D. (1968). *Nature* 218:533.
- Robertson H.D. & Lodish H.F. (1970). *Proc. Nat. Acad. Sci. USA* 67:710.
- Robertson H.D. & Jeppesen P.G.N. (1972). *J. Mol. Biol.* 68:417.
- Rogers J., Clarke P. & Salser W. (1979). *Nucleic Acids Res.* 6:3305.
- Sabo D.L., Domingo E., Bandle E.F., Flavell R.A. & Weissmann C. (1977). *J. Mol. Biol.* 111:235.
- Sanger F. & Brownlee G.G. (1967). *Meth. Enzymol. Nucl. Acids* 12A:361.
- Sanger F., Air G., Barrell B.G., Brown N., Coulson A., Fiddes J., Hutchison C., Slocombe P. & Smith M. (1977). *Nature* 265:687.
- Schaffner W. (1977). Ph.D. thesis, University of Zürich.
- Schwyzer M. (1973). Ph.D. thesis, University of Zürich.
- Schwyzer M., Billeter M.A. & Weissmann C. (1972). *Experientia* 28:750.
- Scott D.W. (1965). *Virology* 26:85.
- Senear A.W. & Steitz J.A. (1976). *J. Biol. Chem.* 251:1902.
- Shapira A. & Billeter M.A. (1978). *J. Biol. Chem.* 253:8400.
- Shine J. & Dalgarno L. (1974). *Proc. Nat. Acad. Sci. USA* 71:1342.
- Silverman P.M. (1973a). *Arch. Bioch. and Biophys.* 157:222.
- Silverman P.M. (1973b). *Arch. Bioch. and Biophys.* 157:234.
- Sinha N.K., Fujimura R.K. & Kaesberg P. (1965). *J. Mol. Biol.* 11:84.
- Sinsheimer R.L. (1976). *Annu. Rev. Biochem.* 46:415.
- Sirover M.A. & Loeb L.A. (1977). *J. Biol. Chem.* 252:3605.
- Smith H.O. & Birnstiel M.L. (1976). *Nucleic Acids Res.* 3:2387.
- Southern E.M. (1975). *J. Mol. Biol.* 98:503.
- Southern E.M. (1979). *Meth. Enzymol.* 68:152.

- Spiegelman S., Pace N.R., Mills D.R., Levisohn R., Eikhom T.S., Taylor M.M., Peterson R.L. & Bishop D.H.L. (1968). Cold Spring Harbor Symp. Quant. Biol. 33:101.
- Staples D.H. & Hindley J. (1971). Nature New Biol. 234:211.
- Staples D.H., Hindley J., Billeter M.A. & Weissmann C. (1971). Nature New Biol. 234:202.
- Stavis R.L. & August J.T. (1970). Annu. Rev. Biochem. 39:527.
- Steitz J.A. (1969). Nature 224:957.
- Steitz J.A. (1972). Nature New Biol. 236:71.
- Steitz J.A. (1975). In "RNA Phages" (ed N.D.Zinder), pp. 319-352, CHSL.
- Steitz J.A. & Jakes K. (1975). Proc. Nat. Acad. Sci. USA 72:4734.
- Stoll E., Wilson K., Reiser J. & Weissmann C. (1976). J.Biol.Chem. 252:990.
- Strauss J.H. (1964). J. Mol. Biol. 10:422.
- Sugiyama T., Korant B.D. & Lonberg-Holm K.K. (1972). Annu. Rev. Microbiol. 26:467.
- Sutcliffe J.G. (1978a). Nucleic Acids Res. 5:2721.
- Sutcliffe J.G. (1978b). Proc. Nat. Acad. Sci. USA 75:3737.
- Szoslak J.W., Stiles J.I., Tye B.K., Chiu P., Sherman F. & Wu R. (1979). Meth. Enzymol. 68:419.
- Taniguchi T. & Weissmann C. (1978). J. Mol. Biol. 118:533.
- Taniguchi T., Palmieri M. & Weissmann C. (1978). Nature 274:223.
- Tinoco I., Uhlenbeck O.C. & Levine M.D. (1971). Nature 230:362.
- Tinoco I., Borer P.N., Dengler P., Levine M.D., Uhlenbeck O.C., Crothers M.D. & Gralla J. (1973). Nature New Biol. 246:40.
- Valentine R.C. & Strand M. (1965). Science 148:511.
- Vandenbergh A., Min Jou W. & Fiers W. (1975). Proc. Nat. Acad. Sci. USA 72:2559.
- Vanderkerckhove J., Francq H. & Van Montagu M. (1969). Arch. Int. Physiol. Biochim. 77:175.
- Vanderkerckhove J., Nolf F. & Van Montagu M. (1973). Nature New Biol. 241:102.
- Viñuela E., Algranati I.D. & Ochoa S. (1967). Eur. J. Biochem. 1:3.

- Wahba A.J., Miller M.J., Niveleau A., Landers T., Carmichael G.G., Weber K., Hawley D.A. & Slobin L.I. (1974). *J. Biol. Chem.* 249:3314.
- Ward R., Strand M. & Valentine R.C. (1968). *Biochem. Biophys. Res. Commun.* 30:310.
- Watanabe L., Miyake T., Sakurai T., Shiba T. & Ohno T. (1967). *Proc. Jap. Acad.* 43:204.
- Weber H. (1976). *Biochim. Biophys. Acta* 418:175.
- Weber H., Billeter M.A., Kahane S., Hindley J., Porter A. & Weissmann C. (1972). *Nature New Biol.* 237:166.
- Weber K. & Konigsberg W. (1975). In "RNA Phages" (ed. N.D.Zinder), pp. 51-84, Cold Spring Harbor Laboratory.
- Weiner A.M. & Weber K. (1971). *Nature New Biol.* 234:206.
- Weiner A.M., Platt T. & Weber K. (1972). *J. Biol. Chem.* 247:3242.
- Weiner A.M. & Weber K. (1973). *J. Mol. Biol.* 80:837.
- Weissmann C. & Ochoa S. (1967). *Progr. Nucleic Acid Res. Mol. Biol.* 6:353.
- Weissmann C., Borst P., Burdon R.H., Billeter M.A. & Ochoa S. (1967). *Proc. Nat. Acad. Sci. USA* 51:890.
- Weissmann C., Colthard L. & Libonati M. (1968a). *Biochemistry* 7:865.
- Weissmann C., Feix G. & Slor H. (1968b). *Cold Spring Harbor Symp. Quant Biol.* 33:83.
- Weissmann C., Billeter M.A., Goodman H.M., Hindley J. & Weber H. (1973). *Annu. Rev. Biochem.* 42:303.
- Wensink P.C., Finegan D.J., Donelson J.E. & Hogness D.S. (1974). *Cell* 3:315.
- Zinder N.D. (1975). In "RNA Phages" (ed. N.D.Zinder), Cold Spring Harbor Laboratory.
- Zinder N.D. & Lyons L.B. (1968). *Science* 159:84.
- Zipper P., Kratky O., Herrmann R. & Hohn T. (1970). *Eur. J. Biochem.* 18:1.

ABBREVIATIONS

(d)A, (d)Ap	(deoxy)adenosine-3'-monophosphate
(d)ATP	(deoxy)adenosine-5'-triphosphate
BSA	bovine serum albumin
(d)CTP	(deoxy)cytidine-5'-triphosphate
cpm	counts per minute (radioactive decay)
DEAE	diethylaminoethyl-
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetate
(d)GTP	(deoxy)guanosine-5'-triphosphate
h	hour(s)
M	molar (mol/l)
m	milli- (10^{-3})
u	micro- (10^{-6})
n	nano- (10^{-9})
p	pico- (10^{-12})
f	femto- (10^{-15})
PEG	polyethylene glycol
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
rRNA	ribosomal ribonucleic acid
S	sedimentation coefficient (Svedberg units)
SDS	sodium dodecyl sulfate
SSC	Standard Saline Citrate (see Materials)
T, Tp	thymidine-3'-monophosphate
TTP	thymidine-5'-triphosphate
TNE	Tris-NaCl-EDTA (see Materials)
Tris	Tris(hydroxymethyl)aminomethane
tRNA	transfer ribonucleic acid

U	units (enzyme activity)
UTP	uridine-5'-triphosphate
v/v	volume per volume (percentage)
w/v	weight per volume (percentage)
Y-(number)	Y-buffer (see section 2.2.4.4.)
Z-(number)	Z-buffer (see section 2.2.4.4)

Curriculum vitae

von Philipp Mekler, geboren am 4. August 1949 in Luzern,
Bürger von Zürich.

- 1956-1961 Primarschule in Luzern
- 1961-1969 Kantonsschule Luzern (Gymnasium); Abschluss mit
Matura Typ B.
- 1969-1975 Studium an der Philosophischen Fakultät II der
Universität Zürich (Hauptfach Molekularbiologie);
Abschluss als dipl.Naturwissenschaftler.
- 1975-1980 Doktorarbeit am Institut für Molekularbiologie I
der Universität Zürich unter der Leitung von
Herrn Prof.Dr.M.A.Billeter.

Während meiner Studienzeit bestand ich folgende Prüfungen:

1. Vordiplom, Sommer 1971 (Zoologie, Mathematik)
 2. Vordiplom, Winter 1973 (Organische Chemie, Bio-
chemie)
- Schlussdiplom, Sommer 1975 (Molekularbiologie, Phy-
sikalische Chemie, Mik-
robiologie)

Ich besuchte die Kurse und Vorlesungen folgender Dozenten:

Mathematik/Physik: Batschelet, van der Waerden; Meyer, Staub

Anorganische Chemie: Oswald

Organische Chemie: Dreiding, Eugster, Schmid, Viscontini, von
Philipsborn

Biochemie: Carafoli, Christen, Humbel, Kägi, Leuthard, Semenza

Physikalische Chemie: Fischer, Labhardt, Wagnière

Zoologie: Burla, Chen, Hadorn, Tardent

Botanik: Hohl, Wanner

Molekularbiologie/Mikrobiologie: Billeter, Birnstiel, Clarkson,
Hennig, Melli, Schaffner, Weber, Weissmann,
Wüthrich; Bachofen, Knüsel

Während meiner Doktorarbeit war ich als Assistent am Institut für
Molekularbiologie I der Universität Zürich angestellt.

